THREE ESSAYS IN SPATIAL ECONOMETRICS

BY

MONALISA SEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Anil Bera, Chair and Director of Research
Professor Geoffrey J.D. Hewings
Professor Daniel McMillen
Professor Xiaofeng Shao

ABSTRACT

A focus on location and spatial interaction has recently gained a more central place not only in applied econometrics but also in theoretical econometrics. The standard econometric techniques often fail in the presence of spatial autocorrelation, which is commonplace in geographic (cross-sectional) data sets, and thus giving misleading inference, and thereby wrong policy implications are derived from these econometric models. In this dissertation I have dealt with such model specification issues which arises due to spatial nature of the data. The main contributions of this dissertation are thus two folds: spatial model specification and specification tests for spatial panel data models. Each chapter provides econometric methods along with empirical examples to demonstrate the importance and utility of the proposed methods.

In Chapter 2, I propose an estimation strategy for popular spatial weight matrix. The spatial lag dependence in a regression model is similar to the inclusion of a serially autoregressive term for the dependent variable in time- series context. However, unlike in the time series model, the implied covariance structure matrix from the spatial autoregressive model can have a very counterintuitive and improbable structure. However, if the weight matrix can capture the underlying dependence structure of the observations then this unintuitive behavior of implied correlation gets corrected to a large extent. Thus in Chapter 2, I explore the possibility of constructing the weight matrix (or the overall spatial dependence in the data) that is consistent with the underlying correlation structure of the dependent variable.

Specification of a model is one of the most fundamental problems in econometrics. However, in most cases, specification tests are carried out in a piecemeal fashion, for example, testing the presence of one-effect at a time ignoring the potential presence of other forms of misspecification. In Chapter 3, I overcome these difficulties by proposing adjusted RS tests

for the panel spatial models under a very general framework, and my proposed test statistics are robust under multiple-forms of misspecification. Most of the existing procedures like likelihood ratio tests (LR) and conditional Lagrange multiplier (LM) tests require estimation of the nuisance parameters. In this respect, a very attractive feature of my approach is that the adjusted tests are based on the joint null hypothesis (of no misspecification) as each the proposed tests take care of the possible presence of all the nuisance parameters through their respective Fisher-Rao score evaluated under joint null and thus requiring estimation of the simplest model.

In Chapter 4, I develop on the theoretical foundation of Chapter 3, by proposing the size-robust tests for dynamic panel models with dynamic space-time dependence. I propose the test statistics robust under local misspecification for time dynamics, individual effects, serial correlation of errors and spatial dependencies like spatial lag and error, and time-space dynamics under the dynamic panel model. Using these proposed tests I investigate the salient features of the data that truly matters for growth analyses. In growth theory different kinds of econometric models have been proposed based on economic theory and the subjective beliefs of researchers, - including simple cross-sectional regression models, panel data models, time series models and recently many types of spatial models. Unfortunately the estimate of growth convergence rate under these different model frameworks vary wildly, even when the same dataset is used. Thus, the question becomes: which model is most appropriate? I propose to address this problem by developing six adjusted Rao's score (RS) tests that are robust under misspecification for a very general dynamic panel model. I start with a simple panel model and then using my proposed test statistics I check whether particular departures (like time dynamics, serial correlation, individual effects, spatial/cross sectional dependence) from this initial specification are supported or rejected by the data. Thus Chapter 4 contributes both to the econometric methodology of specification tests and also tackles with the empirical question of growth convergence debate.

To Thakur, Ma, Swamiji and Gurumaharaj.

ACKNOWLEDGMENTS

I am grateful to my advisor, Professor Anil K. Bera, for his continuous encouragement, advice and generous support in all the projects. I am also thankful to my committee members Professors Geoffrey Hewings, Daniel McMillen, and Xiaofeng Shao for their fruitful advice at the crucial stages of my work. In addition, I am grateful to Professor Hewings and his REAL family for allowing me to use their office space and accepting me as a part of their family too. My thanks also goes to my department, my family and friends for their continuous support and belief in me. Finally, I would like to thank my parents, Dr. Shibani Sen and Pravash K. Sen, to whom I owe everything.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The concept of spatial dependence, or its weaker expression, spatial autocorrelation differs from the more familiar concept of serial correlation in time series domain. Contrasting spatial econometrics to standard econometrics, a narrow definition is offered as dealing with "the specific spatial aspects of data and models in regional science that preclude a straightforward application of standard econometric methods" (Anselin 1988). In this context, spatial dependence is viewed as a special case of cross-sectional dependence, in the sense that the structure of the correlation or covariance between random variables at different locations is derived from a specific ordering, determined by the relative position (distance, spatial arrangement) of the observations in geographic space (or, in general, in network space). While similar to correlation in the time domain, the distinct nature of spatial dependence requires a specialized set of techniques. Importantly, these are not a straightforward extension of time series methods to two dimensions.

Paelinck and Klaassen (1979) specifies five important principles to guide the formulation of spatial econometric models. The five "rules" consist of: (i) the role of spatial interdependence; (ii) the asymmetry in spatial relations; (iii) the importance of explanatory factors located in other spaces; (iv) differentiation between ex post and ex ante interaction; and (v) the explicit modeling of space (topology) in spatial models. In sum, these important dimensions can be identified that define the scope of modern spatial econometric methodology: model specification, estimation, specification testing and spatial prediction. This dissertation mainly contributes to two of these four aspects. The second chapter mainly deals with model specification, whereas, the third and fourth chapter deal with specification tests of these models. All the detail mathematical derivations related to specification tests can be found in the appendices.

The second chapter deals with the spatial model specification issues. Model specification deals with the formal mathematical expression for spatial dependence in regression models. For spatial dependence, this typically takes the form of including spatially lagged variables, i.e., weighted averages of observations for the "neighbors" of a given location. An important aspect of this is the definition of what is meant by neighbors, typically carried out through specification of a spatial weights matrix. Spatially lagged variables can be included for the dependent variable (leading to so-called spatial lag models), explanatory variables (spatial cross-regressive models) and error terms (spatial error models), as well as combinations of these, yielding a rich array of spatially explicit models. Thus, spatial lag dependence in a regression model is similar to the inclusion of a serially autoregressive term for the dependent variable in time- series context. However, unlike in the time series model, the implied covariance structure matrix from the spatial autoregressive model can have a very counterintuitive and improbable structure. A single value of spatial auto correlation parameter can imply a large band of values of pair-wise correlations among different observations of the dependent variable, when the weight matrix for the spatial model is specified exogenously. I illustrate this using cigarette sales data (1963-92) of 46 US states. I observe that two "close" neighbors can have very low implied correlations compared to distant neighbors when the weighting scheme is the first-order contiguity matrix. However, if the weight matrix can capture the underlying dependence structure of the observations then this unintuitive behavior of implied correlation gets corrected to a large extent. Keeping this in mind, in the first chapter of the thesis I explore the possibility of constructing the weight matrix (or the overall spatial dependence in the data) that is consistent with the underlying correlation structure of the dependent variable. The results using my suggested procedure are very encouraging.

The third chapter deals with misspecification tests of spatial panel models. Specification of a model is one of the most fundamental problems in econometrics. In most cases, specification tests are carried out in a piecemeal fashion, for example, testing the presence of one-effect at a time ignoring the potential presence of other forms of misspecification. Many of the suggested tests in literature require estimation of complex models and even then those tests cannot take account of multiple forms of departures. Using Bera and Yoon (1993) general test principle and a spatial panel model framework, we first propose an overall test for “all” possible misspecification. We derive adjusted Rao’s Score (RS) tests for random effect, serial correlation, spatial lag and spatial error, which can identify the definite cause(s) of rejection of the basic model and thus adding in the steps for model revision. For empirical researchers, our suggested procedures provide simple strategies for model specification search using OLS residuals from standard linear model for spatial panel data. Through an extensive simulation study, we evaluate the finite sample performance of our suggested tests and available procedures. We find that the proposed tests have good finite sample properties both in terms of size and power. We also formulate a simple sequential strategy for use in empirical practice.

In the fourth chapter, I have proposed the adjusted Rao’s score (RS) tests that are robust under misspecification for a very general dynamic panel model with cross-sectional dependence and use them for specification search of growth models. In growth theory different kinds of econometric models have been proposed based on economic theory and the subjective beliefs of researchers, - including simple cross-sectional regression, panel data, time series and more recently many types of spatial models. Unfortunately the estimate of growth convergence rate under these different model frameworks vary wildly, even when the same dataset is used. Thus, the question is which model is most appropriate? I use my proposed tests to address this problem and conduct the specification search in multiple directions to understand the underlying data generating process (DGP). Unlike the available tests, these proposed test statistics unravel the interrelation/dependencies among the model parameters and thus make themselves amenable for analysis of misspecification, which is concept-wise similar to analysis of variance. I use the data of 91 non-oil countries over a period of 35 years (1961- 1995) from the Penn World Table, for the specification search. Using the proposed test statistics, I find that heterogeneity, time dynamics and indirect cross-sectional dependence contribute most to the total misspecification than other forms of departures from a simple panel model for this dataset. A very elegant feature of my proposed tests is that they do not require estimation of nuisance parameters, unlike existing tests. Thus the proposed test statistics can identify the underlying DGP without any apriori complex estimation. The extensive simulation study show good finite sample performance of my proposed tests in contrast to other existing procedures. The formulation of these test statistics are quite general and are applicable to many other econometric models for specification search.

CHAPTER 2

THE IMPROBABLE NATURE OF THE IMPLIED CORRELATION MATRIX FROM SPATIAL REGRESSION MODELS

2.1 Introduction

The key idea of modeling of spatial data is that a set of locations can characterize the dependence among the observations. One of the many general ways to do this is to define a neighborhood structure based on the shape of lattice. Among others, are measuring the distance between centroids of the regions. Once this spatial dependence structure is determined or assumed based on distance (social/economic/physical) or adjacency, models resembling time series autoregressive structures are considered. The two very popular models that take into account such spatial dependence structure into account are simultaneously autoregressive (SAR) and conditionally autoregressive (CAR) models. The SAR and CAR models were originally developed by Whittle (1954) and Besag (1974), respectively, mainly on the doubly infinite regular lattice. On regular lattice these models resemble the well understood stationary time series model defined on the integers. On irregular lattice, however, which is most common in economic applications, the effect that the exogenously defined arbitrary neighborhood structure and spatial correlation parameter have on implied covariance structure is not well understood. Wall (2004) was probably the first to do a systematic analysis of the impractical nature of the correlation structure implied by the SAR and CAR models, and this issue has spurred some further inquires, see for instance Martellosio (2009).

In this paper we highlight the problem of implied structure of the SAR model in case of irregular lattice and suggest a possible solution. Although our proposal is for the SAR model, it can be easily extended to the CAR model. Section 2.2 provides a summary of the existing literature. In Section 2.3, we set up the notations and discuss the implied correlation problem arising from the SAR model. Section 2.4 presents empirical example using cigarette sales data on 46 US states, and highlights the unintuitive and impractical behavior of the implied correlation structure when the usual neighborhood matrix is used. Our findings match with the results of Wall (2004). Section 2.5 first gives the basic idea behind our W matrix construction and then we estimate W using Levenberg-Marquardt non-linear optimization procedure. In Section 2.6, we demonstrate how our W matrix helps to correct the implied correlation structure and gives a more intuitive result using the same dataset as in Section 2.4. Section 2.7 concludes the paper. All the figures and tables can be referred from Section 2.8.

2.2 Literature Review

Although the implied correlation structures of the spatial models have such peculiar pattern, it is quite surprising that this issue has received relatively little attention in the literature, given that these models are so widely used in a variety of applications. Haining (1990) and Besag and Kooperberg (1995) mentioned resulting heteroscedasticity from the SAR model with homoscedastic error term. They also pointed out about the unequal covariance between regions that are at same distance apart. The very first systematic treatment of this problem was probably done by Wall (2004). She provided a detailed description of the implied structure of SAR and CAR models, and in particular, considered the dependence and covariance structures on an irregular lattice. Using the US state level summary data of SAT verbal score for the year 1999, she investigated the relationship between the correlation parameter ρ and the implied pairwise correlations among the scores of various states when W was based on first-order neighbors. The implied spatial correlations between the different states using the SAR and CAR models did not seem to follow an intuitive or practical scheme. For example, Wall (2004) found that for the SAR model Missouri and Tennessee are constrained to be the least spatially correlated states, than Tennessee and Arkansas, although all of them are first-order neighbors. Martellosio (2009) shed some further light on how correlation structure of the SAR model depends on W and ρ. He showed that implied correlation between two spatial units depends on particular type of walks connecting the units. When $|\rho|$ is small, the correlation is largely determined by short walks; however, for large values of $|\rho|$, longer walks have more importance. Since ρ can be estimated only after W has been chosen, one cannot control the correlation properties by specifying W. Defining W based on graph, his work explains the inconsistency of ranking of implied correlation between pair of locations as ρ changes and also how the sign of correlation depends on the length of the shortest walk (in graph theoretic sense) from one location to another.

2.3 The SAR Model and the Implied Correlation Problem

Let $y(A_i) : A_i \in (A_1 \ldots A_n)$ be a Gaussian random process where$(A_1 \ldots A_n)$ are n different locations. The value of the variable y in location A_i depends on the values in its neighboring locations A_j. One way to model this dependence is by the simultaneous autoregressive (SAR) model:

$$y = \rho W y + X\beta + \epsilon \tag{2.1}$$

where y is a $n \times 1$ vector observation on the dependent variable, ρ is the spatial autoregressive parameter, $W = ((w_i j))$ is $n \times n$ spatial weight matrix representing degree of potential interactions between neighboring locations (geographic/economic/social), X is $n \times k$ matrix of observations on the explanatory (exogenous) variables, β is $k \times 1$ vector of regression coefficients and ϵ is a $n \times 1$ vector of error term with $\epsilon \sim (0, \sigma^2 I_n)$. Spatial effects are incorporated using the row standardized weight matrix W. One common way to do this is to define $W = (w_i j)$ is

$$w_{ij} = \begin{cases} 1 & \text{if } A_i \text{ shares a common edge or border with region } A_j,\ i \neq j \\ 0 & \text{if } i = j, \\ 0 & \text{otherwise .} \end{cases}$$

The other ways to define the neighborhood structure W is to express weights as functions of the distance between two points or as functions of length of borders. For ease of interpretation, the weight matrix is often standardized such that the elements of each row sum to one. This ensures that all the weights are between 0 and 1, and facilitates the interpretation of operations with the weight matrix as an averaging of neighborhood values. It also ensures that the spatial parameters of different models are comparable. This is not intuitively obvious, but relates to constraints imposed in a maximum likelihood estimation framework, specifically the spatial autocorrelation parameter ρ must lie in the interval $\frac{1}{\omega_{min}}$ to $\frac{1}{\omega_{max}}$, where ω_{min} and ω_{max} are, respectively the smallest and largest eigen values of W [Cliff and Ord (1980)]. For a row standardized matrix, the largest eigen value is +1, and this facilitates the interpretation of ρ as *correlation coefficient*. It is easy to see that the implied covariance matrix of y for model (1) is given by

$$Var(y) = \sigma^2(I - \rho W)^{-1}(I - \rho W)'^{-1} \tag{2.2}$$

Using (2.2), we can calculate the pair-wise correlations $corr(y_i, y_j) = \rho_{ij}$, $i, j = 1, 2, .., n, i \neq j$. However, given ρ and W, these implied ρ_{ij} values can be very hard to interpret in a practical way. To illustrate the implied correlation problem, we first provide some analytical results under two extreme cases of weight matrix.

Case I: All units are neighbors of each other: Cases which may be consistent with this are the ones in which all cross sectional units interact in a confined space. Such a matrix was considered by Case (1992) in a panel data study of the adoption of new technologies by farmers, and by Lee (1999) in a study of the properties of least squares estimators in linear spatial models, and also by Kelijian and Prucha (2002) to evaluate the properties of 2SLS and OLS estimators of SAR models. Here

$$w_{ij} = \begin{cases} \frac{1}{n-1} & \text{for } i \neq j, \\ 0 & \text{for } i = j. \end{cases}$$

Therefore, the weight matrix can be expressed as: $W = \frac{1}{n-1}[J - I],\ n > 1$, where J is the $n \times n$ matrix of ones. It can be verified that $(I - \rho W)^{-1} = \delta_1 J + \delta_2 I$, where $\delta_1 = \frac{\rho}{(n-1+\rho)(1-\rho)}$ and $\delta_2 = \frac{n-1}{n-1+\rho}$. Assuming $\sigma^2 = 1$ we obtain

$$Var(y) = \sigma^2(I - \rho W)^{-1}(I - \rho W)'^{-1} = (n\delta_1^2 + 2\delta_1\delta_2)J + \delta_2^2 I \tag{2.3}$$

Given this variance-covariance structure, it can be seen that correlation matrix goes to I

matrix as $n \to \infty$. Thus, when each unit is neighbor of each other, in the limit the correlation matrix does not depend on ρ !

Case II: Here each unit has only two neighbors. For instance, when $n = 4$ we have

$$W = \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$

which is a tridiagonal Toeplitz form. The implied correlation matrix (using (2)) is also tridiagonal Toeplitz, and is given by

$$Corr(y) = \begin{bmatrix} 1 & \frac{\beta}{\alpha} & \frac{\gamma}{\alpha} & \frac{\beta}{\alpha} \\ \frac{\beta}{\alpha} & 1 & \frac{\beta}{\alpha} & \frac{\gamma}{\alpha} \\ \frac{\gamma}{\alpha} & \frac{\beta}{\alpha} & 1 & \frac{\beta}{\alpha} \\ \frac{\beta}{\alpha} & \frac{\gamma}{\alpha} & \frac{\beta}{\alpha} & 1 \end{bmatrix}$$

where $\alpha = \frac{(2a^2-1)^2+2a^2+4a^4}{(4a^2-1)^2}$, $\beta = \frac{2(2a^2-1)+4a^4}{a(4a^2-1)^2}$, $\gamma = \frac{-4(2a^2-1)_2a^4}{a^2(4a^2-1)^2}$ and $a = -0.5\rho$.

Each element of the inverse of such tridiagonal matrix is non zero (El-Shehaway, El-Shreff, Al-Henaway (2008)). Here units 1 and 3 are not connected ($w_{13} = w_{31} = 0$) directly, but we have a non zero implied correlation. In spatial context it implies that even if two units are "not" neighbors of each other, they can have very high non - zero implied spatial correlations. This can be interpreted as the spill-over effects from neighbors.

These examples are somewhat artificial. Therefore in the next section, using the widely applied cigarette sales data on 46 States, we demonstrate that a single value of ρ can imply a large band of values of ρ_{ij} with the same w_{ij} values. Our findings confirm the results of Wall (2004).

2.4 An Empirical Example

In order to analyze the spatial interaction and implied correlation structure of a SAR model we consider the 1963 - 1992 cigarette sales data on 46 states, that has been widely used for panel data analysis by Baltagi and Levin (1992) and Baltagi, Griffin and Xiong (2000), and later by Elhorst (2005) for spatial panel analysis. The underlying model is:

$$log(C) = \alpha + \rho W log(C) + \beta_1 log(P) + \beta_2 log(Y) + \beta_3 log(P_n) + \epsilon \tag{2.4}$$

where C is real per capita sales of cigarettes to persons of smoking age (14 years and older), measured in packs of cigarettes per capita, P is the average retail price of a pack of cigarettes measured in real terms, Y is the real per capita disposable income, and P_n denotes the minimum real price of cigarettes in any neighboring state. This last variable is a proxy for the smuggling effect across state borders, and acts as a substitute price attracting consumers from high-tax states to cross over to low-tax states. As in Elhorst (2005), we use the conventional form of row-standardized first-order neighborhood weight matrix, and in Table 2.1, present the estimation results based on 1992 cross-section data for the 46 states.

To illustrate the behavior of the implied correlation structure from the estimated SAR model, in Figure 2.1, we display the histogram of all the implied first-order neighbor correlations and notice a wide variation. In order to check if we find the similar pattern, we also did the same analysis with Columbus crime data (Anselin (1998)). In this dataset The crime variable (CRIME) pertains to the (1980) combined total of residential burglaries and vehicle thefts per households in the of 49 contiguous census track neighborhood of Columbus, Ohio. Figure2.2 shows that we indeed have similar pattern as in Figure 2.1.

The smallest correlation for Cigarette sales data is 0.09 that occurs between Missouri and Tennessee and the largest correlation, equal to 0.37, occurs between New Hampshire and Maine. Wall (2004) also noted smallest and largest implied correlations exactly for these states, though she used different data (1999 US statewide average SAT verbal scores) and model. The common feature between Wall's and our situations is the W matrix, more specifically, Maine has only one neighbor, i.e., New Hampshire, and Tennessee and Missouri have 7 and 8 neighbors, respectively. Also the qualitative nature of the histograms of Wall (in her Figure 3 with $\hat{\rho} = 0.60$) and ours are very similar. Therefore, we can say that implied correlation is simply a function of the first-order neighbors each state has. To elaborate further on the implied correlations of Missouri and Tennessee with their 8 and 7 neighbors, respectively, from Table 2.2, we note that Missouri is more correlated with Kansas than with Tennessee; and Tennessee is more correlated with its neighbor Alabama than with Missouri. Such peculiarity arises mainly due to the nature of covariance matrix (1.2) that involves inversion of the sparse matrix $(I - \rho W)$. Our relative ranking of the states using implied spatial correlation almost coincides with that of Wall (2004). These two datasets have no connection economically; and ranking of implied spatial correlation is determined by the priory fixed weight matrix.

Figure 2.3 demonstrates that the relationship between the implied correlation and number of neighbors is not that simple. If number of neighbors is less, then implied correlation is strong. There is a band in which the implied correlations vary for a given number of neighbors, and we observe less heterogeneity for extreme number of neighbors. This pattern is also observed for Columbus crime data as shown in Figure 2.4.

Now we focus on how implied correlations behave as functions of true parameter ρ (i.e., irrespective of data). From Figure 2.5, we observe that for any given ρ, there is a high variability in correlations. For example, when $\rho = 0.1$, the implied correlations vary from 0.03 to 0.13; while for $\rho = 0.6$, they vary from 0.25 to 0.73. As ρ increases the implied correlations of all locations increases monotonically, which matches the behavior of autoregressive models in time series, i.e., correlation increases with the autoregressive parameter. However, the most unintuitive behavior is that as ρ changes, there are many lines that cross each other, implying the inconsistency of ranking of implied correlations. For example, when $\rho = 0.2$ the correlation (Missouri, Arkansas) = 0.17 and correlation (Tennessee, Arkansas) = 0.24. However, when $\rho = 0.7$, correlation (Missouri, Arkansas) = 0.33 and correlation (Tennessee, Arkansas) = 0.26. Wall (2004) reported the same phenomenon. Surprisingly, the same pattern is observed even when using the Columbus crime data. Thus, we find the same unintuitive nature of implied correlation while using a very different dataset too. Figure 2.6 can be compared to Figure 2.5. Therefore, the implied correlations of SAR model with first-order neighbor W matrix do exhibit some unintuitive and impractical behavior.

2.5 Numerical Optimization

It is a general understanding that the weight matrix captures the *spatial-links* of the observations as Ord (1975) stated that the (i, j)th element of W *represents the degree of possible interaction of location* j *on location* i. However, each element of $(I - \rho W)^{-1}(I - \rho W)'^{-1}$ provides the correlation structure of y. As evident from Wall (2004) and from our illustration above, if one expresses the spatial dependence in terms of neighborhood matrix W, then the covariance from $(I - \rho W)^{-1}(I - \rho W)'^{-1}$ does not have a logical connection to the spatial correlation.The choice of spatial weights is a central component of spatial models as it imposes a priori structure on spatial dependence. Although the existing literature contains an implicit acknowledgment of the issues of choosing an appropriate weight matrix, most empirical studies treat W known, fixed and arbitrary spatial weight matrix (Giacomini and Granger 2004). We propose to construct the weight matrix using past time series data to remove the odd features of implied correlations discussed above.

Suppose the dependent variable y_i is observed over n locations, where $i = 1, \dots, n$ for $t = 1, \dots, T$ in *past* T periods. We estimate the variance covariance matrix $V(y) = \sum$, whose (i,j)th element is given by $\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \overline{y}_i)(y_{jt} - \overline{y}_j)$, where $y_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $y_j = \frac{1}{T}\sum_{t=1}^{T} y_{jt}$. Our objective is to investigate the implied correlation structure of a SAR model at the current period, and therefore, construction of the weight matrix based on past observations helps us to avoid the endogeneity issue. We solve the following system for W.

$$\sigma^2(I - \rho W)^{-1}(I - \rho W)'^{-1} = \sum$$

We can take $\sigma^2 = 1$, which will have no consequence for our solution to W. Also, since W is row standardized, the solution will be invariant to ρ. Therefore, without loss of generality we solve

$$(I - \rho W)^{-1}(I - \rho W)'^{-1} = \sum$$

i.e.,

$$(WW') - (W + W') = \sum^{-1} - I. \tag{2.5}$$

We need to find W that solves the equation (2.5) subject to

i) $w_{ii} = 0$

ii) $w_{ij} \geq 0$

iii) $\sum_j w_{ij} = 1$.

Alternatively, our objective is to find a solution to the constrained system of nonlinear equations:

$$F(w) = I + (w \times w') - (w + w') - \sum^{-1} = 0, \quad w \in W, \tag{2.6}$$

where$W \subseteq \mathcal{R}^{m+}$ is a nonempty, closed and convex set and $F : O \to \mathcal{R}^m$ is a given mapping defined on an open neighborhood O of the set W. Here $m = n^2$, where n is the number of locations. We denote by W^* the set of solutions to (2.5). To solve (2.6) we minimize: $f(w) = ||F(w)||^2$, where $||.||$ is the Euclidean norm, subject to the constraints as above. We employ Levenberg-Marqaurdt (LM) algorithm that interpolates between Gauss-Newton algorithm and method of gradient descent. In many cases, LM algorithm is more robust than Gauss-Newton as it finds a solution even if it starts very far off from the optimal values. It is an iterative procedure where in each step w is replaced by $w + d$. To determine d, the function $F(w + d)$ are approximated by their linearization using Taylor Theorem i.e., $F(w + d) \approx F(w) + J \times d$, where $J = \frac{\partial F(w)}{\partial w}$ is the gradient of F with respect to w. At its minimum, the gradient of f with respect to d will be zero. The above 1st order approximation gives

$$f(w + d) \approx ||F(w + d)||^2 \approx ||F(w) + J \times d||^2.$$

Taking derivative with respect to d and setting the result equal to zero gives $(J^T J)d = -J^T F(w)$, where J is the Jacobian term. This gives us a set of linear equations which can be solved for the increment vector d. Levenberg-Marquardt contribution is to replace this equation by a *damped version*,

$$(J^T J + \mu \times diag(J^T J))d = -J^T F(w).$$

The main difference between Gauss-Newton and LM algorithm is in terms of normal equations. In LM algorithm the normal equations are modified in such a way that the increment vector d is always rotated towards the direction of steepest descent. In a more formal way, LM type method for this system of equations generates a sequence w^k by setting $w^{k+1} = (w^k + d^k)$, where d^k is the solution to the linearized subproblem:

$$min\ \theta^k(d) = ||F(w^k) + J_k d||^2 + \mu_k ||d||^2, \quad s.t \quad w^k + d \in W. \tag{2.7}$$

Here, J_k is an approximation of Jacobian of $F'(w^k)$ and μ_k is the positive parameter. Note that θ^k is a strictly convex quadratic function, hence the solution d^k of (2.6) always exists uniquely. Since our constraints is of box constraints type, any iterate w^k can be projected easily into the feasible region W. The feasible region of W is such that any $w \in W$ has the structure defined by the above constraints. Therefore, we set $w^{k+1} = P_W(w^k + d_u^k), k = 0, 1, \ldots,$ where P_W is the projection matrix and d_u^k is the unique solution to the unconstrained subproblem:

$$min\ \theta^k(d_u), d_u \in \mathcal{R}^m.$$

We call this projected LM method since the unconstrained step gets projected onto the feasible region W. The projected version of LM algorithm needs significantly less time per iteration since the strict convexity of the function θ^k ensures that d_u^k is a global minimum of this function if and only if $\bigtriangledown \theta^k(d_u^k) = 0$, i.e., if and only if d_u^k is the unique solution of the system of linear equations [for detailed discussion on Levenberg- Marquardt Method, see Nocedal and Wright (2006)]:

$$(J_k^T J_k + \mu_k diag(J_k^T J_k))d_u = -J_k^T F(w^k) \tag{2.8}$$

The step-by-step algorithm is as follows:

S1) Choose $w^0 \in W, \mu > 0, v > 1, \gamma > 0$ and set $k = 0$, tolerance=1e-10.

S2) If $F(w^k) < tolerance$, then Stop, otherwise go to S3.

S3) Compute $J_k = F'(w^k)$.

S4) Set $\mu_k = \frac{\mu}{v^k}$ and compute d_u^k as a solution to (8).

S5) If $||F(P_W(w^k + d_u^k)|| \leq \gamma||F(w^k)||$, then set $w^{k+1} = P_W(w^k + d_u^k)$, update k to k+1 and go to S2; Otherwise go to S6.

S6) Set $\mu_k = \mu \times v^k$ and compute d_u^k as a solution to (8).

S7) If $||F(P_W(w^k + d_u^k)|| \leq \gamma||F(w^k)||$, then set $w^{k+1} = P_W(w^k + d_u^k)$, update k to $k+1$ and go to S2.

Note, if any kth iteration comes to S6, then for $k + 1th$ iteration onwards, it will flow as $S2 \rightarrow S3 \rightarrow S6 \rightarrow S7$. This is due to the choice of dampening factor as suggested by Marquardt (1963). If there is no reduction in residual by setting $\mu_k = \frac{\mu}{v^k}$, then the dampening factor is increased by successive multiplication by v until a better point is found with the new dampening factor $\mu_k = \frac{\mu}{v^k}$ for some k. However, if the use of $\mu_k = \frac{\mu}{v^k}$ results in reduction of residuals then this is taken as a new value of μ and the process continues. In other words, as μ_k gets small, the algorithm approaches the Gauss-Newton algorithm, if μ_k becomes large with successive iterations, it approaches the steepest gradient algorithm. The technique invented by Levenberg-Marquardt involves *blending* between these two extremes. It uses a steepest descent type method until our objective function approaches a minimum, and then gradually switches to the quadratic rule. It tries to guess how close we are to a minimum by how our error is changing. The intuition is simple; i.e., if error is increasing, then our quadratic approximation is not working well and we are likely not near a minimum, so we should increase μ_k in order to blend more towards simple gradient descent. Conversely, if error is decreasing, our approximation is working well, and we expect that we are getting closer to a minimum so μ_k is decreased to bank more on the Hessian. The algorithm we used is very similar to the projected LM algorithm of Kanzow-Yamashita-Fukushima (2002). As long as F is affine and twice continuously differentiable, any accumulation point of the sequence w^k generated by our algorithm, is a stationary point of (2.7).

2.6 Application of the Proposed Solutions

We estimate the SAR model (2.4) for the year 1992 using our proposed weight matrix. In order to avoid endogeneity problem, we construct our W matrix using the data on C (Cigarette sales) from 46 states for the period 1963 - 1991. Table 2.3 provides the estimates of the SAR model using the standard W matrix and our numerically solved W using Levenberg-Marquardt algorithm. It is clear that the estimated SAR model using our proposed W matrix is equally good compared to that with the standard W in terms of log-likelihood value.

In Figure 2.7 we plot the first-order implied correlation as a function of weights from our estimated W. Out of $46 \times 46 = 2116$ pairs of locations, we only plots the 188 first-order neighbor correlations. We first arrange the weights of 188 pairs of first-order neighbors in ascending order, and then the implied correlations are sorted out in ascending order as well. From the Figure 2.7 we note that the implied correlations have very slow increasing trend with weights. Also there is little variation. This is in contrast to Figure 2.3 (where number of neighbor increases means weight decreases) which displayed much higher variation.
In contrast to Figure 2.5 now for each value of ρ, the band of variation of implied correlations is very narrow in Figure 2.8. For example, when $\rho = 0.1$, the implied correlations vary only in between 0.004 and 0.006; while for $\rho = 0.6$ they vary from 0.09 to 0.11. Also now there is no crossing, and thus the inconsistency of the ranking of implied correlations seen in Figure 2.3, is absent in Figure 2.8.

Finally, to address the implied heterogeneity of SAR model, in Figure 2.9, we plot the 46 diagonal elements of $\sum$ as a function of the number of first-order neighbors. Using the first-order contiguity matrix leads to substantial variation of implied variances of y_i(which decreases with the number of neighbors). In contrast, our proposed W matrix hardly produces any implied heterogeneity.

2.7 Conclusion

We first demonstrate that the unintuitive and impractical nature of the implied correlations implied by the estimated SAR models with row standardized neighborhood matrix.

We propose a simple methodology for estimation of spatial weight matrix. Our procedure yields very intuitive results in terms of implied correlations and variances. Our proposed methodology is illustrated using the cigarette sales data. Although we apply our proposed method only to the SAR model, it can be easily extended to the CAR model. For CAR, $Var(y) = \sigma^2(I - \rho W)^{-1}$, which is a variation of (2.5), and therefore, one can apply the LM procedure to construct a more meaningful weight matrix.

2.8 Tables and Figures

Table 2.1: Estimation Results of Model (2.4) (Standard errors are in parentheses)

Parameters	Ordinary Least Squares	Spatial Autoregressive Model
$\hat{\beta}_1$	-1.24 (0.31)	-1.15 (0.29)
$\hat{\beta}_2$	-1.17 (0.32)	0.27 (0.30)
$\hat{\beta}_3$	1.03 (0.19)	0.74(0.15)
$\hat{\rho}$	N/A	0 .28 (0.14)
$\hat{\sigma^2}$	0.05	0.04
Log-Lik		25.78
R^2	0.15	0.18

Table 2.2: Implied Correlation Between First- Order Neighbors of Missouri and Tennessee

Missouri		Tennessee	
1st order neighbors	Implied correlation	1st order neighbors	Implied correlation
Arkansas	0.0965	Alabama	**0.1354**
Illinois	0.1062	Arkansas	0.1036
Iowa	0.0977	Georgia	0.1256
Kansas	**0.1516**	Kentucky	0.0931
Kentucky	0.0879	Mississippi	0.1325
Nebraska	0.1108	Missouri	**0.0873**
Oklahoma	0.1110	Virginia	0.1044
Tennessee	**0.0873**		

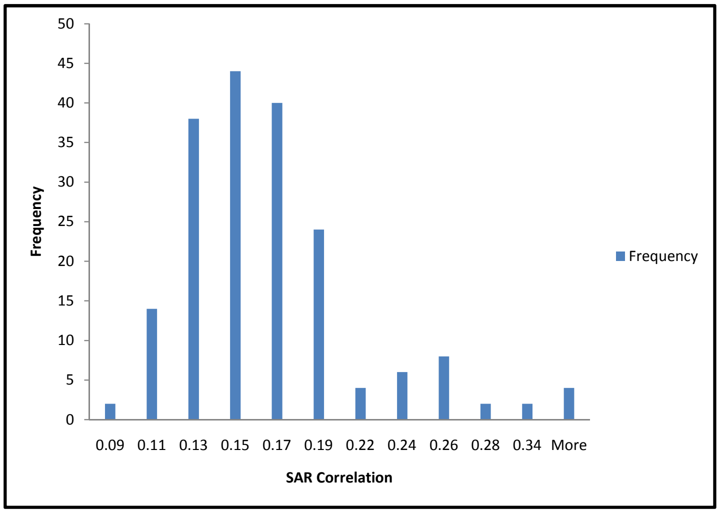


Figure 2.1: Histogram Of Implied Correlations for Cigarette Sales data

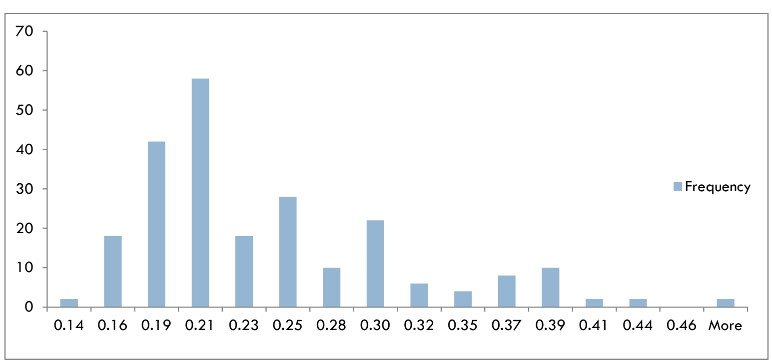


Figure 2.2: Histogram Of Implied Correlations for Columbus Crime data

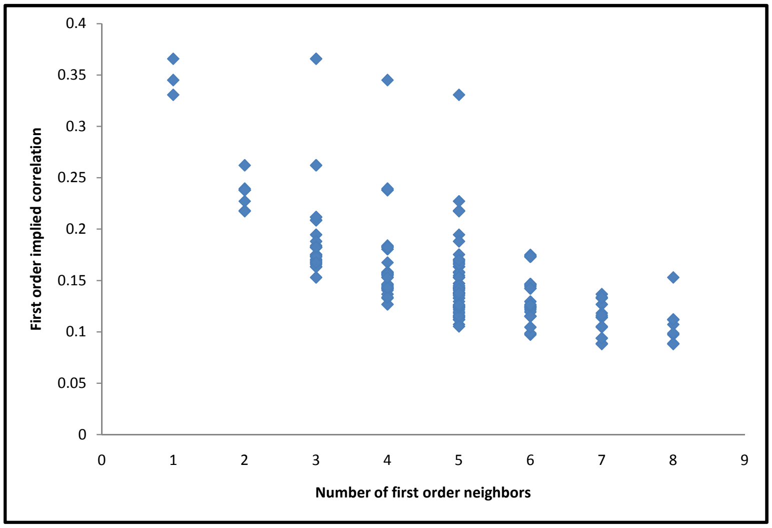


Figure 2.3: Scatter Plot of Implied Correlations Of SAR Model for Cigarette Sales data

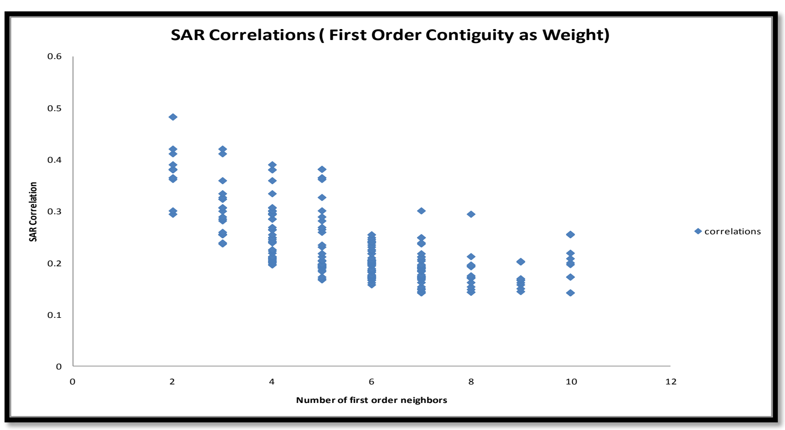


Figure 2.4: Scatter Plot of Implied Correlations Of SAR Model for Columbus Crime data

Table 2.3: Estimation Results of Model (2.4) (Standard errors are in parentheses)

Parameters	SAR (Proposed W)	SAR (1st order neighborhood W)
$\hat{\beta}_1$	-1.10 (0.29)	-1.15 (0.29)
$\hat{\beta}_2$	-0.18 (0.29)	0.27 (0.30)
$\hat{\beta}_3$	0.55 (0.17)	0.74(0.15)
$\hat{\rho}$	0.45 (0.16)	0 .28 (0.14)
$\hat{\sigma^2}$	0.03	0.04
Log-Lik	26.37	25.78
R^2	0.27	0.18

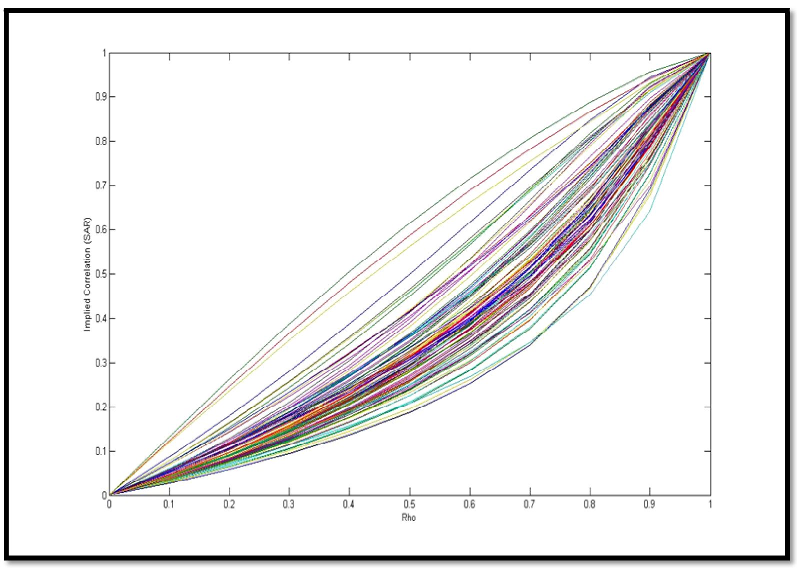


Figure 2.5: Implied Correlations Of SAR Model (as a function of ρ) for Cigarette sales data

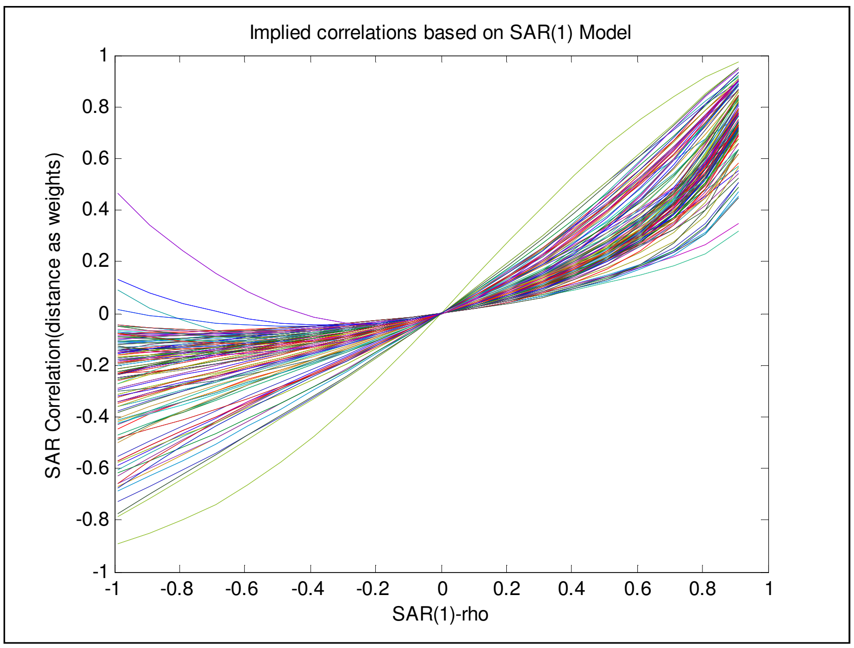


Figure 2.6: Implied Correlations Of SAR Model (as a function of ρ) for Columbus crime data

Figure 2.7: Implied Correlations Of SAR Model (W=Constructed)

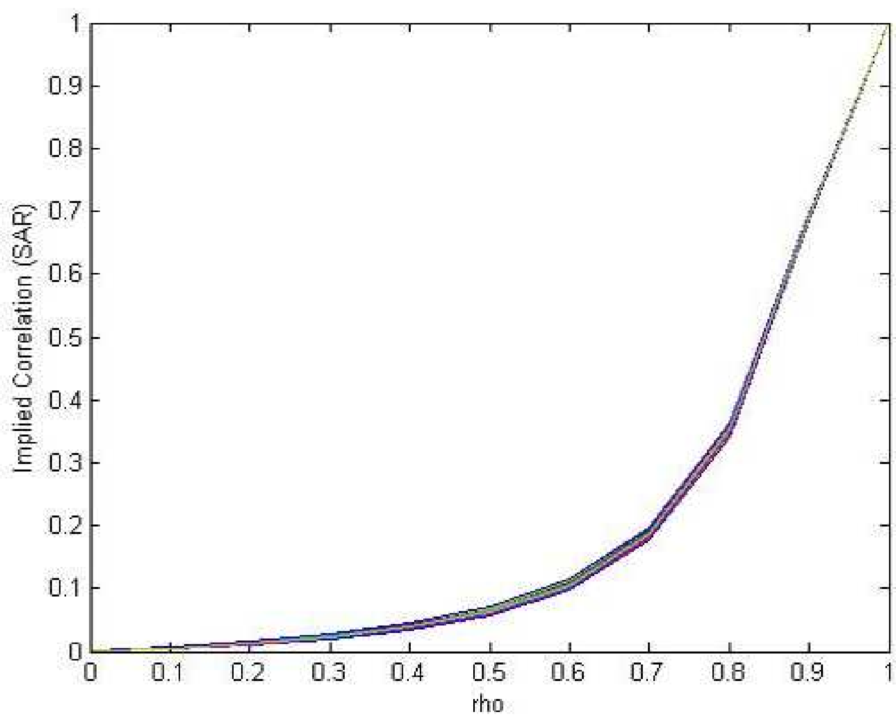


Figure 2.8: Implied Correlations Of SAR Model (as a function of rho)

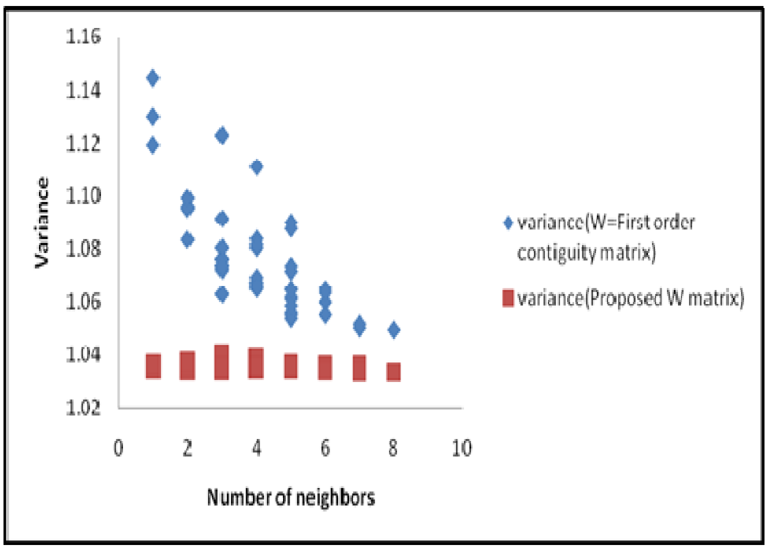


Figure 2.9: Variance Comparison Of 46 US States

CHAPTER 3

SPECIFICATION TESTING FOR PANEL SPATIAL MODEL

3.1 Introduction

Econometricians' interest on problems that arise when the assumed model (used in constructing a test) deviates from the data generating process (DGP) goes a long way back. As emphasized by Haavelmo (1944), in testing any economic relations, specification of a set of possible alternatives, called the *priori admissible hypothesis*, Ω^0, is of fundamental importance. Misspecification of the priori admissible hypotheses was termed as type-III error by Bera and Yoon (1993), and Welsh (1996, p. 119) also pointed out a similar concept in the statistics literature. Broadly speaking, the alternative hypothesis may be misspecified in three different ways. In the first one, what we shall call "complete misspecification", the set of assumed alternative hypothesis, Ω^0, and the DGP $\Omega^{'}$, say, are mutually exclusive. This happens, for instance, if in the context of panel data model, one test for serial independence when the DGP has random individual effects but no serial dependence. The second case, "underspecification" occurs when the alternative is a subset of a more general model representing the DGP, i.e., $\Omega^0 \subset \Omega^{'}$. This happens, for example, when *both* serial correlation and individual effects are present, but are tested separately (one at a time assuming absence of other effect). The last case is "overtesting" which results from overspecification, i.e., when $\Omega^0 \supset \Omega^{'}$. This can happen if a joint test for serial correlation and random individual effects is conducted when only one effect is present in DGP. It can be expected that consequences of overtesting may not be that serious (possibly will only lead to some loss of power), whereas those of undertesting can lead to highly misleading results, seriously affecting both size and power [See Bera and Jarque (1982) and Bera (2000)]. Using the asymptotic distributions of standard Rao's score (RS) test under local misspecification, Bera and Yoon (1993) suggested an adjusted RS test that is robust under misspecification and asymptotically equivalent to the optimal Neyman's $C(\alpha)$ test. As we will discuss, an attractive feature of this approach is that the adjusted test is based on the joint null hypothesis of no misspecification, thereby requiring estimation of the simplest model. A surprising additivity property also enables us to calculate the adjusted tests quite effortlessly.

The plan of the rest of the paper is as follows. In the next section we provide a brief review of existing literature. Section 3.3 gives the main results on the general theory of RS tests when the alternative is misspecified, and presents the modified RS test which is robust under local misspecification. We develop the spatial panel model framework in Section 3.4 and present the log-likelihood function. Section 3.5 formulates the new diagnostic tests which take account of misspecification in multiple directions. To illustrate the usefulness of our proposed tests, in Section 3.6, we demonstrate how our methodology can assist a practitioner to reformulate his/her model using an empirical example. In particular, we use Heston, Summers and Aten (2002) Penn World Table, which contains information on real income, investment and population (among many other variables) for a large number of countries and the growth-model proposed by Ertur and Koch (2007). In Section 3.7 we provide evidence of good finite sample performance of our suggested and some available tests based on an extensive simulation study. Section 3.8 concludes the paper. All the tables and figures are relegated to Section 3.9.

3.2 A Brief Survey of the Literature

The origins of specification testing for spatial models can be traced back to Moran (1950a, 1950b). Much later this area was further enriched by many researchers, for example, see Cliff and Ord (1972), Brandsma and Ketellapa (1979a, 1979b), Burridge (1980), Anselin (1980, 1988a, 1988b, 1988c) and Kelejian and Robinson (1992). Most of these papers focused on tests for specific alternative hypothesis in the form of either spatial lag or spatial error dependence based on ordinary least squares (OLS) residuals. As we discussed above, their separate applications when other or both kinds of dependencies are present will lead to unreliable inference. It may be natural to consider a *joint* test for lag and error autocorrelations. Apart from the problem of overtesting (when only one kind of dependence characterizes the DGP) mentioned above, the problem with such a test is that we cannot identify the exact nature of spatial dependence once the joint null hypothesis is rejected. One approach to deal with this problem to use conditional tests, i.e., to use test for spatial error dependence after estimating a spatial lag model, and vice versa. This, however, requires maximum likelihood (ML) estimation, and the simplicity of test based on OLS residuals is lost. Anselin, Bera, Florax and Yoon (1996) was possibly the first paper to study systematically the consequences of testing one kind of dependence (lag or error) at a time. Using the Bera and Yoon (1993) general approach, Anselin et al. (1996) developed OLS-based adjusted RS test for lag (error dependence) in the possible presence of error (lag) dependence. Their Monte Carlo study demonstrated that the adjusted tests are very capable of identifying the exact source(s) of dependence and they have very good finite sample size and power properties. In a similar fashion, in context of panel data model, Bera, Sosa-Escudero and Yoon (2001) showed that when one tests for either random effects or serial correlation without taking account of the presence of other effect, the test rejects the true null hypothesis far too often under the presence of the unconsidered parameter. They found that the presence of serial correlation made the Bruesh and Pagan (1980) test for random effects to have excessive size. Similar over rejection occur for the test of serial correlation when the presence of random effect is ignored. Bera et al. (2001) developed size-robust tests (for random effect and serial correlation) that allow distinguishing the source(s) of misspecification in specific direction(s).

Now if we combine the models considered in Anselin et al. (1996) and Bera et al. (2001), we have the *spatial panel model*, potentially with *four* sources of departure (from the classical regression model) coming from four extra parameters: the spatial lag, spatial error, random effect and (time series) serial correlation parameters. The spatial panel model has been studied extensively and has gained much popularity over time given the wide availability of the longitudinal data. (See, for instance, Elhorst (2003), Pesaran (2004), Lee (2009), Pesaran and Tosseti (2011)). In this paper, we investigate a number of strategies to test against *multiple* form of misspecification of this kind. Using a general model we derive an overall test and a number of adjusted tests that take the account of possible misspecification in multiple directions. For empirical researchers our suggested procedures provide simple strategies to identify specific direction(s) in which the basic model needs revision using only OLS residuals from the standard linear model for spatial panel data. As we further discuss below, all the available tests in the literature take into account of only *one* or *two* potential sources of misspecification at a time, and many of them require ML estimation to account for the nuisance parameter(s).

Recently, many researchers have conducted conditional and marginal specification tests in spatial panel data models. Baltagi, Song and Koh (2003) proposed conditional LM tests, which test for random regional effects given the presence of spatial error correlation and also, spatial error correlation given the presence of random regional effects. Baltagi et al. (2007) adds another dimension to the correlation in the error structure, namely, serial correlation in the remainder error term. Both these were based on the extension of spatial error models (SEM). Baltagi and Liu (2008) developed similar LM and LR tests with spatial lag dependence and random individual effects in a panel data regression model. Their paper derives conditional LM tests for the absence of random individual effects without ignoring the possible presence of spatial lag dependence and vice-versa. Baltagi, Song and Kwon (2009) considered a panel data regression with heteroscedasticity as well as spatially correlated disturbances. As in previous works, Baltagi et al. (2009) derived the conditional LM and marginal LM tests. However the specification tests proposed in the above papers require ML estimation of nuisance parameters and such a strategy will complex as we add more parameters to generalize the model in multiple directions. Recently, based on Bera and Yoon (1993) (henceforth BY), Montes-Rojas (2010) has proposed an adjusted RS test for autocorrelation in presence of random effects and vice-versa, after estimating the spatial dependent parameter using MLE and instrumental variable estimation methods. The possibility of using estimator to construct RS-type adjusted tests means wide applicability of the robustness approach that we are using. Since our basic model can be estimated efficiently by OLS, we do not have to resort to any other estimation.

3.3 A General Approach to Testing in the Presence of Nuisance Parameters

Consider a general model represented by the log-likelihood $L(\gamma, \psi, \phi)$ where the parameters γ, ψ and ϕ are, respectively, $(p \times 1), (r \times 1)$ and $(s \times 1)$ vectors. Here we will assume that underlying density function satisfies the regularity conditions, as stated in Surfling [pp], Lehman and Romano [pp], for the MLE to have asymptotic Gaussian distribution. Suppose a researcher sets $\phi = \phi_0$ and tests $H_0 : \psi = \psi_0$ using the log-likelihood function $L_1(\gamma, \psi) = L(\gamma, \psi, \phi_0)$, where ψ_0 and ϕ_0 are known. The Rao score statistic for testing H_0 in $L_1(\gamma, \psi)$ will be denoted by RS_ψ. Let us denote $\theta = (\gamma', \psi', \phi')_{(p+r+s)x1}$ and $\tilde{\theta} = (\tilde{\gamma'}, \psi_0', \phi_0')_{(p+r+s)\times 1}$, where $\tilde{\gamma}$ is the ML estimator of γ under $\psi = \psi_0$ and $\phi = \phi_0$. We define the score vector and the information matrix, respectively, as

$$d_a(\theta) = \frac{\partial L(\theta)}{\partial a}$$

for $a = \gamma, \phi$ and ψ and

$$J(\theta) = -E[\frac{1}{NT}\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}] = \begin{bmatrix} J_{\gamma_{p\times p}} & J_{\gamma\psi_{p\times r}} & J_{\gamma\phi_{p\times s}} \\ J_{\psi\gamma_{r\times p}} & J_{\psi_{r\times r}} & J_{\psi\phi_{r\times s}} \\ J_{\phi\gamma_{s\times p}} & J_{\phi\psi_{s\times r}} & J_{\phi_{s\times s}} \end{bmatrix} \tag{3.1}$$

where n denotes the sample size. If $L_1(\gamma, \psi)$ were the true model, then it is well known that under $H_0 : \psi = \psi_0$,

$$RS_\psi = \frac{1}{N} d_\psi(\tilde{\theta})' J_{\psi.\gamma}^{-1}(\tilde{\theta}) d_\psi(\tilde{\theta}) \to^D \chi_r^2(0) \tag{3.2}$$

where $\to^D$ denotes convergence in distribution and

$$J_{\psi.\gamma}(\theta) = J_\psi - J_{\psi\gamma} J_\gamma^{-1} J_{\gamma\psi} \tag{3.3}$$

Under local alternatives $H_1 : \psi = \psi_0 + \frac{\epsilon}{\sqrt{n}},\;\; RS_\psi \to^D \chi_r^2(\lambda_1)$, where the non centrality parameter $\lambda_1 = \epsilon' J_{\psi.\gamma}\epsilon$. Given this set-up, asymptotically the test will have correct size and will be locally optimal [see Bera and Bilias (2001a)]. Now suppose that the true log-likelihood function is $L_2(\gamma, \phi)$ so that the considered alternative $L_1(\gamma, \psi)$ is (completely) misspecified. Using the local misspecification $\phi = \phi_0 + \frac{\delta}{\sqrt{n}}$, Davidson and MacKinnon (1987) and Saikkonen (1989) derived the asymptotic distribution of RS_ψ under $L_2(\gamma, \phi)$ as $RS_\psi \to^D \chi_r^2(\lambda_2)$, where the non-centrality parameter $\lambda_2 = \delta' J_{\phi\psi.\gamma} J_{\psi.\gamma}^{-1} J_{\psi\phi.\gamma}\delta$ with

$$J_{\phi\psi.\gamma} = J_{\psi\phi} - J_{\psi\gamma} J_\gamma^{-1} J_{\gamma\phi} \tag{3.4}$$

Owing to the presence of this non-centrality parameter λ_2, RS_ψ will reject the true null hypothesis $H_0 : \psi = \psi_0$ more often, i.e., the test will have excessive size. Here the crucial term is $J_{\psi\phi.\gamma}$ [see equation (3.4)] which can be interpreted as the partial covariance between the score vectors d_ψ and d_ϕ after eliminating the linear effect of d_γ on d_ψ and d_ϕ. If $J_{\psi\phi.\gamma} = 0$, then asymptotically the local presence of the parameter ϕ has no effect on RS_ψ. Using (3.4), BY suggested a modification to RS_ψ so that the resulting test is valid in the local presence of ϕ. The modified statistic is given by

$$RS_\psi^* = \frac{1}{N}[d_\psi - J_{\psi\phi.\gamma} J_{\phi.\gamma}^{-1} d_\phi]'[J_{\psi.\gamma} - J_{\psi\phi.\gamma} J_{\phi.\gamma}^{-1} J_{\phi\psi.\gamma}]^{-1}[d_\psi - J_{\psi\phi.\gamma} J_{\phi.\gamma}^{-1} d_\phi] \tag{3.5}$$

This new test essentially adjusts the mean and variance of the standard RS_ψ. Another way to look at RS_ψ^* is to view the quantity$J_{\psi\phi.\gamma} J_{\phi.\gamma}^{-1} d_\phi$ as the prediction of d_ψ by d_ϕ, and thus $d_\psi - J_{\psi\phi.\gamma} J_{\phi.\gamma}^{-1} d_\phi = d_{\psi.\phi}^*$ say, is the part of d_ψ that remains after eliminating the effect of d_ϕ. In the literature, $d_{\psi.\phi}^*$ is known as the effective score of ψ, which is orthogonal to d_ϕ [see Bera and Bilias (2001b)]. Under $\psi = \psi_0$ and RS_ψ^* has a central χ_r^2 distribution. Thus, under misspecification RS_ψ^* has the same asymptotic null distribution central χ_r^2 as that of RS_ψ with $\psi = \psi_0$ and $\phi = \phi_0$, thereby producing an asymptotically correct size test even when the model is locally misspecified. Under the local alternatives $\psi = \psi_0 + \frac{\epsilon}{\sqrt{n}}$

$$RS_\psi^* \to \chi_1^2(\lambda_3) \tag{3.6}$$

where $\lambda_3 = \epsilon'(J_{\psi.\gamma} - J_{\psi\phi.\gamma} J_{\phi.\gamma}^{-1} J_{\phi\psi.\gamma})\epsilon$. Note that $\lambda_1 - \lambda_3 \geq 0$, where λ_1 is given in (1). Result (2.6) is valid both in presence or absence of the local misspecification $\phi = \phi_0 + \frac{\delta}{\sqrt{n}}$, since the asymptotic distribution of RS^*_ψ is unaffected by the local departure of ϕ from ϕ_0. Therefore, RS^*_ψ will be less powerful than RS_ψ when there is no misspecification. The quantity

$$\lambda_1 - \lambda_3 = \epsilon'(J_{\psi\phi.\gamma}J^{-1}_{\phi.\gamma}J_{\phi\psi.\gamma})\epsilon \tag{3.7}$$

can be regarded as the premium we pay for the validity of RS^*_ψ under local misspecification i.e. the cost for robustness. BY further show that for local misspecification the adjusted test is asymptotically equivalent to Neyman's $C(\alpha)$ test and thus shares its optimal properties. Three observations are worth noting regardingRS^*_ψ. First, RS^*_ψ requires estimation only under the joint null, namely $\psi = \psi_0$ and $\phi = \phi_0$. That means, in most cases, as we will see later, we can conduct our tests based on only OLS residuals. Given the full specification of the model $L(\gamma, \psi, \phi)$, it is of course possible to derive RS test for $\psi = \psi_0$ after estimating ϕ (and γ) by MLE, which are generally referred to as conditional tests. However, ML estimation of ϕ could be difficult in some instances. Second, when $J_{\psi\phi.\gamma} = 0$, $RS^*_\psi = RS_\psi$, which is a simple condition to check, as we indicated earlier. If this condition is true, RS_ψ [in equation (2)] is an asymptotically valid test in the local presence of ϕ. Finally, let $RS_{\psi\phi}$ denote the joint RS test statistic for testing hypothesis of the form $H_0 : \psi = \psi_0$ and $\phi = \phi_0$ using the alternative model $L(\gamma, \psi, \phi)$. Then it be shown that [for a proof see Bera, Bilias and Yoon (2007), Bera, Montes-Rojas and Sosa-Escudero (2009)]

$$RS_{\psi\phi} = RS^*_\psi + RS_\phi = RS_\psi + RS^*_\phi \tag{3.8}$$

where RS_ϕ and RS^*_ϕ are, respectively, the counterparts of RS_ψ and RS^*_ψ for testing $H_0 : \phi = \phi_0$. This is a very important identity since it implies that a joint RS test for two parameter vectors ψ and ϕ can be decomposed into sum of two orthogonal components: (i) the adjusted statistic for one parameter vector and (ii) (unadjusted) marginal test statistic for the other. Since many econometrics softwares provide the marginal (and sometime the joint) test statistics, the adjusted versions can be obtained effortlessly. In the context of spatial panel model which will be introduced in the next section, ψ and ϕ will denote any combinations of the four parameters relating spatial lag, spatial error, random effect and serial correlation, and the parameter vector γ will correspond to the basic regression model.

Test statistic in (3.6) will be based on the OLS estimator $\tilde{\gamma}$ under the joint null $H_0 : \psi = \psi_0$ and $\phi = \phi_0$, i.e., absence of any kind of correlations and random effect. In a way, $RS_{\psi\phi}$ can be viewed as the total departure from the joint null hypothesis of no misspecification in the basic model. When the true model is presented by $L(\gamma, \psi, \phi)$, $RS_{\psi\phi}$ is "always" a valid test, and under local alternatives $\psi = \psi_0 + \frac{\epsilon}{\sqrt{n}}$ and $\phi = \phi_0 + \frac{\delta}{\sqrt{n}}$ [See Bera et al. (2001)]

$$RS_{\psi\phi} \rightarrow^D \chi^2_{r+s}(\lambda_4)$$

where, $\lambda_4 = [\epsilon' \quad \delta'] \begin{bmatrix} J_{\psi.\gamma} & J_{\psi\phi.\gamma} \\ J_{\phi\psi.\gamma} & J_{\phi.\gamma} \end{bmatrix} \begin{bmatrix} \epsilon \\ \delta \end{bmatrix}$

Significance of $RS_{\psi\phi}$ indicates some form of misspecification in the basic model that involves parameter γ. However, we can identify the correct source(s) of departure only by using the adjusted statistics (RS^*_ψ and RS^*_ϕ) not the marginal ones (RS_ψ and RS_ϕ). Our testing strategy is close to the idea of Hillier (1991) in the sense that we try to partition an overall rejection region to obtain evidence about the specific direction(s) in which the basic model needs revision. And we achieve that without estimating any of the nuisance parameters. As we will discuss later, the Fisher-Rao score functions evaluated even under the joint null can be viewed as the sufficient statistics of the underlying parameters. [see Bera and Bilias (2001b)] Thus heuristically, the RS test statistic contains as much as information about the total misspecification in the basic model but requires very little estimation.

3.4 A Spatial Panel Model

We consider the following spatial panel model:

$$y_{it} = \tau \sum_{j=1}^{N} m_{ij} y_{jt} + X_{it}\beta + u_{it} \tag{3.9}$$

$$u_{it} = \mu_i + \epsilon_{it} \tag{3.10}$$

$$\epsilon_{it} = \lambda \sum_{j=1}^{N} w_{ij}\epsilon_{jt} + v_{it} \tag{3.11}$$

$$v_{it} = \rho v_{it-1} + e_{it}, \text{where } e_{it} \sim IIDN(0, \sigma_e^2) \tag{3.12}$$

for $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$. Here y_{it} is the observation for i^{th} location/unit at t^{th} time, X_{it} denotes the observations on non-stochastic regressors and e_{it} is the regression disturbance. Spatial dependence is captured by the weight matrices $M = (m_{ij})$ and $W = (w_{ij})$. Here m_{ij} and w_{ij} are the (i, j) th element of weight matrices M and W respectively, which capture the interdependence of income and unobserved error terms between the country i and j.The matrices M and W are each row-standardized and the diagonal elements are set to zero. In this model framework, time dynamics (γ), random effects (μ_i) with $\mu_i \sim IID(0, \sigma_\mu)$, serial correlation (ρ), space recursive (δ), spatial lag dependence (τ) and spatial error dependence (λ) are considered.

In matrix form, the equations (3.9) - (3.12) can be written compactly as

$$y = \tau(I_T \otimes M)y + X\beta + u, \tag{3.13}$$

where y is of dimension $NT \times 1, X$ is $NT \times K, \beta$ is $k \times 1$, u is $NT \times 1$, I_T is an identity matrix of dimension $T \times T$ and $\otimes$ denotes Kronecker product. Here X is assumed to be of full column rank and its elements are bounded in absolute value. The disturbance term can be expressed as

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})v. \tag{3.14}$$

Here $B = (I_N - \lambda W)$ and ι_T is vector of ones of dimension T. Under this setup, the variance-covariance matrix of u is given by

$$\Omega = \sigma_{\mu^2}[J_T \otimes I_N] + [V \otimes (B'B)^{-1}], \tag{3.15}$$

where J_T is a matrix of ones of dimension $T \times T$, and V is the familiar $T \times T$ variance -covariance matrix for AR (1) process in equation (3.12), i.e.,

$$V = E(v'v) = [\frac{1}{1-\rho^2}V_1] \otimes \sigma_e^2 I_N = V_\rho \otimes \sigma_e^2 I_N, \tag{3.16}$$

with

$$V_1 = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & 1 \end{bmatrix},$$

and $V_\rho = \frac{1}{1-\rho^2} V_1$.

The log-likelihood function of the above model can be written as:

$$L = \frac{-NT}{2} ln 2\pi - \frac{1}{2}\ln|\Omega| + T\ln|A| - \frac{1}{2}[(I_T \otimes A)y - X\beta]'\Omega^{-1}[(I_T \otimes A)y - X\beta] \quad (3.17)$$

where $A = (I_N - \tau M)$. Following Baltagi et. al (2007), I can write

$$\frac{1}{2}\ln|\Omega| = -\frac{N}{2}\ln(1-\rho^2) + \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| + \frac{NT}{2}\ln\sigma_e^2 - (T-1)\ln|B|,$$

where $d^2 = \alpha^2 + (T-1)$, $\alpha = \sqrt{\frac{1+\rho}{1-\rho}}$ and $\phi = \frac{\sigma_\mu^2}{\sigma_e^2}$. Substituting $\frac{1}{2}\ln|\Omega|$ in L , I obtain

$$L = \frac{-NT}{2} ln 2\pi + \frac{N}{2}\ln(1-\rho^2) - \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| - \frac{NT}{2}\ln\sigma_e^2 + (T-1)\ln|B| + T\ln|A| \\ - \frac{1}{2}[(I_T \otimes A)y - X\beta]'\Omega^{-1}[(I_T \otimes A)y - X\beta] \quad (3.18)$$

3.5 Derivation of the Specification Tests

We are interested in testing $H_0 : \psi = 0$ in the possible presence of the parameter vector ϕ. For the spatial panel model the full parameter vector is given by $\theta = (\beta, \sigma_e^2, \sigma_\mu^2, \rho, \lambda, \tau)'$. In context of our earlier notation $\theta = (\gamma', \psi', \phi')'$, $\gamma = (\beta', \sigma_e^2)$ and ψ and ϕ could be any combinations of the parameters under test, namely $(\sigma_\mu^2, \rho, \lambda, \tau)$. The main advantage of using RS test principal is that we need estimation of θ_0 only under the joint null $H_0^a : \sigma_\mu^2 = \rho = \lambda = \tau = 0$ i.e., of $\theta' = (\beta', \sigma_e^2, 0, 0, 0, 0)'$. For simplicity we assume the weight matrices W and M to be same. This is often realistic in practice, since there may be good reasons to expect the structure of spatial dependence to be the same for the dependent variable Y and the disturbance term ϵ. On the basis of the derivations given in the Appendix, the score functions and the information matrix J evaluated under H_0^a i.e., restricted MLE of θ_0 with

$\tilde{\gamma} = (\tilde{\beta}, \tilde{\sigma_e^2})'$ are :

$$\frac{\partial L}{\partial \beta} = 0 \tag{3.19}$$

$$\frac{\partial L}{\partial \sigma_e^2} = 0 \tag{3.20}$$

$$\frac{\partial L}{\partial \sigma_\mu^2} = \frac{NT}{2\tilde{\sigma_e^2}}[\frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1] \tag{3.21}$$

$$\frac{\partial L}{\partial \rho} = \frac{NT}{2}[\frac{\tilde{u}'(G \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{3.22}$$

$$\frac{\partial L}{\partial \tau} = \frac{\tilde{u}'[(I_T \otimes W)Y_{NT}]}{\tilde{\sigma_e^2}} \tag{3.23}$$

$$\frac{\partial L}{\partial \lambda} = \frac{NT}{2}[\frac{\tilde{u}'(I_T \otimes (W + W'))\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{3.24}$$

where $\tilde{u} = y - x\tilde{\beta}$ is the OLS residual vector of dimension $NT \times 1$, $\tilde{\sigma_e^2} = \frac{\tilde{u}'\tilde{u}}{NT}$ and $G = \frac{\partial V_1}{\partial \rho}|_{H_o^a}$, where G is bidiagonal matrix with bidiagonal elements all equal to one. Y_{NT} is vector of y of dimension $NT \times 1$.

The information matrix J, equation (3.1), under H_o^a is

$$J(\theta_0) = \begin{bmatrix} \frac{X'X}{\tilde{\sigma_e^2}} & 0 & 0 & 0 & 0 & \frac{X'(I_T \otimes W)X\tilde{\beta}}{\tilde{\sigma_e^2}} \\ 0 & \frac{NT}{2\tilde{\sigma_e^4}} & \frac{NT}{2\tilde{\sigma_e^4}} & 0 & 0 & 0 \\ 0 & \frac{NT}{2\tilde{\sigma_e^4}} & \frac{NT}{2\tilde{\sigma_e^4}} & \frac{N(T-1)}{\tilde{\sigma_e^2}} & 0 & 0 \\ 0 & 0 & \frac{N(T-1)}{\tilde{\sigma_e^2}} & N(T-1) & 0 & 0 \\ 0 & 0 & 0 & 0 & Ttr(W^2 + WW') & Ttr(W^2 + WW') \\ \frac{X'(I_T \otimes W)X\tilde{\beta}}{\tilde{\sigma_e^2}} & 0 & 0 & 0 & Ttr(W^2 + WW') & H \end{bmatrix}$$

where $J = E(-\frac{1}{NT}\frac{\partial^2 L}{\partial\tilde{\theta}\partial\tilde{\theta}'})$ evaluated at θ_0. Here $H = Ttr(W^2 + WW') + \frac{\tilde{\beta}'X'(I_T \otimes W')(I_T \otimes W)X\tilde{\beta}}{\tilde{\sigma_e^2}}$. The detailed derivation and expression of each of the terms of the information matrix are relegated to the appendix A.

Apart from the RS statistic for full joint null hypothesis H_0^a, I propose *four* (modified) test statistics for the following hypotheses:

I) $H_o^b : \sigma_\mu^2 = 0$ in presence of ρ, λ, τ.

II) $H_o^c : \rho = 0$ in presence of $\sigma_\mu^2, \lambda, \tau$.

III) $H_o^d : \lambda = 0$ in presence of σ_μ^2, ρ, τ.

IV) $H_o^e : \tau = 0$ in presence of $\sigma_\mu^2, \rho, \lambda$.

These four will guide us to identify the correct source(s) of departure(s) from H_0^a when it is rejected. One can test various combinations of I) to IV) by testing two/three parameters at a time under the null and compute additional ten test statistic (as is done sometimes in practice). However, we would argue that is not necessary. Also keeping the total number of tests to a minimum is beneficial to avoid the pre-testing problem since in practice, researchers reformulate their model based on test outcomes. There is a big advantage in considering test statistics that require estimation only under the joint null. Given the full specification of the model in equation (3.9) - (3.12), it is of course possible to derive conditional RS and likelihood ratio (LR) tests, for say, $\sigma_\mu^2 = 0$ in the presence of ρ, λ, τ as advocated in Baltagi et al. (2003), Baltagi et al. (2007) and Baltagi and Liu (2008). However, that requires ML estimation of (ρ, λ, τ) (and also of σ_μ^2 for LR test).

Let us take the case I, i.e., for $H_o^b : \sigma_\mu^2 = 0$ in presence of ρ, λ, τ, the term $J_{\psi\phi.\gamma}$ i.e., $J_{\sigma_\mu^2\phi.\beta\sigma_e^2} \neq 0$ where $\psi = \sigma_\mu^2$ and $\phi = (\rho, \lambda, \tau)$. Thus the parameter σ_μ^2 is not "independent" of (ρ, λ, τ) and vice-versa and therefore, the marginal RS test statistic based on the raw score $d_{\sigma_\mu^2}$, i.e. $RS_{\sigma_\mu^2}$ for $H_o^b : \sigma_\mu^2 = 0$ assuming $\phi = (\rho, \lambda, \tau)' = (0, 0, 0)$ is not a valid test under the presence of ρ, λ, τ. Instead $RS^*_{\sigma_\mu^2}$ which eliminates the effects of (ρ, λ, τ) without estimating them, would be a more appropriate statistic, as discussed above. Therefore, the focus of our strategy is to carry out the specification test for a general model with minimum estimation. As we will see later from our Monte Carlo results, we lose very little in terms of finite sample size and power. Though $RS^*_{\sigma_\mu^2}$ does not require explicit estimation of (ρ, λ, τ), effect of these parameters have been taken into account through the use of the effective score $d^*_{\sigma_\mu^2.\phi}$. Of course, given the current computing power, it is not that difficult to estimate a complex model. However, it could be sometime hard to ensure the stability of many parameter estimates. Also theoretically the stationarity regions of the parameter space have not been fully worked out as discussed in Elhorst (2010).

We now discuss the test statistics for each of the above hypotheses. From equation (3.5) recall the form of locally size adjusted RS for $H_0 : \psi = 0$ in presence of parameter ϕ:

$$RS^*_\psi = \frac{1}{N}[d_\psi - J_{\psi\phi.\gamma}J^{-1}_{\phi.\gamma}d_\phi]'[J_{\psi.\gamma} - J_{\psi\phi.\gamma}J^{-1}_{\phi.\gamma}J_{\phi\psi.\gamma}]^{-1}[d_\psi - J_{\psi\phi.\gamma}J^{-1}_{\phi.\gamma}d_\phi]$$

For each of the following test, $\gamma = (\beta', \sigma_e^2)$, ψ is one parameter of $(\sigma_{\mu^2}, \rho, \lambda, \tau)$ and ϕ is the rest three.

Let us consider the test statistics one-by-one. Detail derivations are in Appendix A.

I) $H_o^b : \sigma_\mu^2 = 0$ in presence of ρ, λ, τ.

Here we are testing the significance of random location/ individual effect in presence of time series autocorrelation of errors, spatial error dependence and spatial lag dependence.

In particular we have $\psi = \sigma_\mu^2$ and $\phi = (\rho, \lambda, \tau)'$. Here

$$J_{\psi\phi.\gamma} = (J_{\sigma_{\mu^2}\rho}\ 0\ 0) = [\frac{N(T-1)}{\tilde{\sigma_e^2}}\ 0\ 0]$$

which implies that the unadjusted RS is not a valid test under the local presence of ρ, λ, τ . However, note that only the partial covariance between $d_{\sigma_\mu^2}$ and d_ρ is nonzero, while it is zero for $d_{\sigma_\mu^2}$ and d_λ; $d_{\sigma_\mu^2}$ and d_τ . This fact is reflected in the unadjusted and adjusted version of the test statistic for H_o^b:

$$RS_{\sigma_\mu^2} = \frac{(d_{\sigma_\mu^2})^2}{J_{\sigma_{\mu^2}.\sigma_e^2}} = \frac{NTA^2}{2(T-1)}$$

$$RS^*_{\sigma_\mu^2} = \frac{[d_{\sigma_\mu^2} - J_{\sigma_\mu^2\rho}J_\rho^{-1}d_\rho]^2}{J_{\sigma_\mu^2.\sigma_e^2} - J_{\sigma_\mu^2\rho}J_\rho^{-1}J_{\rho\sigma_\mu^2}} = \frac{NT^2(A-B)^2}{2(T-1)(T-2)} \tag{3.25}$$

where $A = \frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1$ and $B = \frac{\tilde{u}'(G \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}}$.

In the numerator of $RS^*_{\sigma_\mu^2}$, the effective (adjusted) score $d^*_{\sigma_\mu^2} = [d_{\sigma_\mu^2} - J_{\sigma_\mu^2\rho}J_\rho^{-1}d_\rho]$ is that part of $d_{\sigma_\mu^2}$ which is "orthogonal" to d_ρ. For other nuisance parameters λ (spatial error lag) and τ (spatial dependence lag) we do not need to make any adjustment since they do not have any asymptotic effect on σ_μ^2 as far as the testing is concerned. Similar interpretation applies to the denominator of $RS^*_{\sigma_\mu^2}$ which reflects the adjustment needed in variance part for changing the raw score to effective score. Thus inference regarding σ_μ^2 is affected only by the presence of ρ and is independent of the spatial aspects of the model. This separation between "time" and "space" aspects panel spatial model is quite interesting, and we consider it a plus point that our adjusted test take account of such information implied by model. Moreover from equations (3.4) and (3.6) we know that overall power (in presence of local misspecification) of $RS_{\sigma_\mu^2}$ and $RS^*_{\sigma_\mu^2}$ are guided by the non-centrality parameter respectively.

$$\lambda_2(\sigma_\mu^2) = \sigma_\mu^4(J_{\sigma_\mu^2\rho}J^{-1}_{\sigma_{\mu^2}.\sigma_e^2}J_{\rho\sigma_\mu^2}) = \frac{N\rho^2}{T}(T-1)$$

$$\lambda_3(\sigma_\mu^2) = \sigma_\mu^4(J_{\sigma_{\mu^2}.\sigma_e^2} - J_{\sigma_\mu^2\rho}J_\rho^{-1}J_{\rho\sigma_\mu^2}) = \frac{N\sigma_\mu^4}{2\sigma_e^4}(T-1)(T-2) \tag{3.26}$$

II) $H_o^c : \rho = 0$ in presence of $\sigma_\mu^2, \lambda, \tau$.

Here we test the significance of time-series autocorrelation in presence of random effect, spatial lag and spatial error dependence effects. For this test we have $\psi = \rho$ and $\phi = (\sigma_\mu^2, \lambda, \tau)'$. Here

$$J_{\psi\phi.\gamma} = (J_{\rho\sigma_{\mu^2}}\ 0\ 0) = [\frac{N(T-1)}{\tilde{\sigma_e^2}}\ 0\ 0]$$

Again this expression can be given similar interpretation as above, i.e., the inference on ρ will be affected only by the presence of random effect, not by the presence of spatial dependence, given β and σ_e^2. The unadjusted and adjusted test statistics for this case are:

$$RS_\rho = \frac{(d_\rho)^2}{J_\rho} = \frac{NT^2B^2}{4(T-1)^2}$$

$$RS^*_\rho = \frac{[d_\rho - J_{\rho\sigma_\mu^2}J^{-1}_{\sigma_{\mu^2}.\sigma_e^2}d_{\sigma_\mu^2}]^2}{J_\rho - J_{\rho\sigma_\mu^2}J^{-1}_{\sigma_\mu^2.\sigma_e^2}J_{\sigma_\mu^2\rho}} = \frac{NT^2(B-\frac{2A}{T})^2}{4(T-1)(T-\frac{2}{T})} \tag{3.27}$$

Here also the overall power of RS_ρ and RS^*_ρ in case of local misspecification can be obtained respectively from

$$\lambda_2(\rho) = \frac{N\sigma_\mu^4}{\sigma_e^4}(T-1)$$

$$\lambda_3(\rho) = \rho^2(J_\rho - J_{\rho\sigma_\mu^2}J^{-1}_{\sigma_\mu^2.\sigma_e^2}J_{\sigma_\mu^2\rho}) = N(T-1)(1-\frac{2}{T}) \tag{3.28}$$

For the rest of the two cases we provide only the algebraic expression:

III) $H_o^d : \lambda = 0$ in presence of σ_μ^2, ρ, τ.

We are testing the significance of spatial error dependence in presence of random individual/location effects, error autocorrelation and spatial lag dependence effect. Here$\psi = \lambda$ and $\phi = (\sigma_\mu^2, \rho, \tau)'$ and

$$J_{\psi\phi.\gamma} = (0 \ 0 \ J_{\lambda\tau})$$

where $J_{\lambda\tau} = Ttr(W^2 + WW')$. The test statistics are

$$RS_\lambda = \frac{d_\lambda^2}{J_\lambda} = \frac{[\frac{NT\tilde{u}'(I_T \otimes (W+W'))\tilde{u}}{2\tilde{u}'\tilde{u}}]^2}{\tilde{T}^2}$$

$$RS_\lambda^* = \frac{[d_\lambda - J_{\lambda\tau}J_{\tau.\beta}^{-1}d_\tau]^2}{J_\lambda - J_{\lambda\tau}J_{\tau.\beta}^{-1}J_{\tau\lambda}} = \frac{ZZ'}{\tilde{T}[1 - \tilde{T}J_{\tau.\beta}^{-1}]} \tag{3.29}$$

where $Z = \frac{1}{2\tilde{\sigma_e^2}}[\tilde{u}'E\tilde{y} - \tilde{T}J_{\tau.\beta}^{-1}(\tilde{u}(E+E')\tilde{u})],\ E = (I_T \otimes W)$ and $\tilde{T} = Ttr(W^2 + WW')$.

Like the above two cases, the non-centrality parameter of RS_λ^* and RS_λ^* are respectively

$$\lambda_2(\lambda) = \tau^2\tilde{T}$$

$$\lambda_3(\lambda) = \lambda^2(J_\lambda - J_{\lambda\tau}J_{\tau.\beta}^{-1}J_{\tau\lambda}) = \lambda^2\tilde{T}[1 - \tilde{T}J_{\tau.\beta}^{-1}] \tag{3.30}$$

IV) $H_o^e : \tau = 0$ in presence of $\sigma_\mu^2, \rho, \lambda$.

Here$\psi = \tau$ and $\phi = (\sigma_\mu^2, \rho, \lambda)'$ and

$$J_{\psi\phi.\gamma} = (0 \ 0 \ J_{\tau\lambda})$$

where $J_{tau\lambda} = Ttr(W^2 + WW')$. The test statistics are

$$RS_\tau = \frac{d_\tau^2}{J_{\tau.\beta}} = \frac{(\tilde{u}'E\tilde{y})^2}{\tilde{\sigma_e^4}J_{\tau.\beta}}$$

$$RS_\tau^* = \frac{[d_\tau - J_{\tau\lambda}J_\lambda^{-1}d_\lambda]^2}{J_{\tau.\beta} - J_{\tau\lambda}J_\lambda^{-1}J_{\lambda\tau}} = \frac{[\frac{1}{2\tilde{\sigma}_e^2}[\tilde{u}'E\tilde{y} - (\tilde{u}(E+E')\tilde{u})]]^2}{J_{\tau.\beta} - \tilde{T}} \tag{3.31}$$

Similarly, the non-centrality parameter of RS_τ and RS_τ^* in presence of nuisance parameters are

$$\lambda_2(\tau) = \lambda^2 J_{\tau.\beta}^{-1}\tilde{T}^2$$

$$\lambda_3(\tau) = \tau^2(J_{\tau.\beta} - J_{\tau\lambda}J_\lambda^{-1}J_{\lambda\tau}) = \tau^2(J_{\tau.\beta} - \tilde{T}) \tag{3.32}$$

Going back to the influential separation between time (non-spatial) dimension of the model (σ_μ^2 and ρ) and its spatial counterpart (λ and τ) we show that the following partial covariances are all zero.

$J_{\rho(\lambda\tau).\sigma_\mu^2\gamma} = 0$ i.e., the partial covariance between d_ρ and $d_\tau.d_\lambda$ after eliminating the effect of $d_{\sigma_\mu^2}$ and d_γ is zero.

Similarly, $J_{\sigma_\mu^2(\lambda\tau).\rho\gamma} = 0$, $J_{\lambda(\sigma_\mu^2\rho).\tau\gamma} = 0$, $J_{\tau(\sigma_\mu^2\rho).\lambda\gamma} = 0$ and $J_{(\lambda\tau)(\sigma_\mu^2\rho).\gamma} = 0$.

By decomposing the joint RS for $H_0^a : \sigma_\mu^2 = \rho = \lambda = \tau = 0$, as in equation (3.6) we obtain

$$RS_{\sigma_\mu^2\rho\lambda\tau} = RS_{\sigma_\mu^2\rho} + RS_{\lambda\tau} = RS_{\sigma_\mu^2}^* + RS_\rho + RS_\lambda^* + RS_\tau = RS_{\sigma_\mu^2} + RS_\rho^* + RS_\lambda + RS_\tau^* \tag{3.33}$$

As expected the omnibus test statistic $RS_{\sigma_\mu^2\rho\lambda\tau}$ is not the sum of four marginal RS statistics. The above result support our finding that the unadjusted RS over rejects the null as it fails to take into account of the effect of the relevant interaction effects within the spatial and non-spatial parameters. From the above decomposition, one can trivially obtain the adjusted RS tests from their unadjusted counterparts as follows:

$$RS_{\sigma_\mu^2}^* = RS_{\sigma_\mu^2\rho} - RS_{\sigma_\mu^2} \tag{3.34}$$

$$RS_\rho^* = RS_{\sigma_\mu^2\rho} - RS_\rho \tag{3.35}$$

$$RS_\lambda^* = RS_{\lambda\tau} - RS_\tau \tag{3.36}$$

$$RS_\tau^* = RS_{\lambda\tau} - RS_\lambda \tag{3.37}$$

This provides a enormous computational simplicity for practitioners. One can easily obtain the joint RS (two directional) and marginal RS (one directional) for the parameters using any popular statistical package like STATA, R, Matlab, based on an OLS residuals, and then obtain the adjusted test statistics as above. Thus our methodology is implementable without any computational burden, unlike the LR and conditional LM tests.

For instance, if $RS_{\sigma^2_\mu \rho}$ is not significant one can accept the hypothesis that $\sigma^2_\mu = \rho = 0$. If on the other hand $RS_{\lambda\tau}$ is significant, then RS^*_λ and RS^*_τ should be used to detect the exact source(s) of rejection. One testing strategy would be to apply $RS_{\sigma^2_\mu \rho\lambda\tau}$ first comparing it with χ^2_4 critical value, at say, α - level. If it is significant, then use $RS_{\sigma^2_\mu \rho}$ and $RS_{\lambda\tau}$ separately employing χ^2_2 say $\frac{\alpha}{2}$ level. Depending on which ones are significant, we can use the adjusted tests, even lower level, say $\frac{\alpha}{4}$ using χ^1_2 critical values. This way one can limit the overall significance level of the battery of tests used.

3.6 An Empirical Illustration

We now present an empirical example that illustrates the usefulness of our proposed tests. The data consist of a sample of 91 countries over the period 1961-1995. These countries are those of the Mankiw, Romer and Weil (1992) non-oil sample, for which Heston, Summers and Aten (2002) Penn World Table (PWT version 6.1) provide data. We use a slight variation of Ertur and Koch (2007)'s growth model that explicitly takes account of technological interdependence among countries and examines the impact of neighborhood effect. The magnitude of physical capital externalities at steady state, which is not usually identified in the literature, is estimated using a spatially augmented Solow model. We illustrate how a practitioner, after estimating the simplest model would proceed to identify the dependent structures and reformulate the model accordingly. We consider the following model which is a slight variation of Ertur and Koch (2007):

$$\ln(\frac{Y_{it}}{L_{it}}) = \beta_0 + \beta_1 \ln s_{it} + \beta_2 \ln(n_{it} + g + \delta) + \gamma_1 \sum_{j \neq i}^{N} w_{ij} \ln(\frac{Y_{jt}}{L_{jt}}) + \sum_{j \neq i}^{N} w_{ij} (\beta_3 \ln s_{jt} + \beta_4 \ln(n_j t + g + \delta)) + u_{it}$$

$$u_{it} = \mu_i + \epsilon_{it}$$

$$\epsilon_{it} = \gamma_2 \sum_{j \neq i}^{N} w_{ij}\epsilon_{jt} + v_{it}$$

$$v_{it} = \rho v_{it-1} + e_{it}$$

where Y is real GDP, L is the number of workers, s is the saving rate, and n is the average growth of the working-age population (ages 15 to 64), The coefficients δ, depreciation of physical capital and g is the balanced growth rate are taken to be known at value 0.05 $(\delta + g)$ as is common in the literature (Mankiw-Romer-Weil (1992), Ertur Koch (2007)]. w_{ij} are the elements of weight matrix W, based on geographical distance as illustrated in Ertur and Koch (2007).

We estimate the model using OLS method under our null hypothesis, i.e., when all the four effects are absent and then compute the following test statistics: (i) joint test for all four departures, i.e. random effect, serial correlation, spatial error lag and spatial lag, ($RS_{\sigma_\mu^2\rho\lambda\tau}$), (ii) joint test for random effect and serial correlation ($RS_{\sigma_\mu^2\rho}$), (iii) joint test for spatial error lag and lag dependence ($RS_{\lambda\tau}$), (iv) the Breush-Pagan test for random effects ($RS_{\sigma_\mu^2}$), (v) the proposed modified version ($RS^*_{\sigma_\mu^2}$), (vi) the RS test of serial correlation test (RS_ρ), (vii) the corresponding modified version (RS^*_ρ), (viii) the RS test of spatial error dependence (RS_λ), (ix) proposed modified version (RS^*_λ) (x) RS test of spatial lag dependence test (RS_τ), and (xi) lastly the derived modified version (RS^*_τ) . To identify specific departure(s) there is no need to consider any other combination of tests due to the asymptotic dependence we discussed earlier and we are considering the unadjusted RS tests mainly for comparison purpose; we identify the one not informative in practical application. The test statistics are presented in Tables 2.1 and 2.2.

All of the test statistics are computed individually, and we verified the equalities in equations (3.29) - (3.33). The omnibus statistic ($RS_{\sigma_\mu^2\rho\lambda\tau} = 220.02$) rejects the joint null when compared to χ_2^4 critical value at any level. Later in our Monte Carlo case study we will demonstrate the good finite sample size of $RS_{\sigma_\mu^2\rho\lambda\tau}$. More specifically its estimated size is 0.059 when the nominal size is 0.05. The joint tests for random effect and serial correlation $RS_{\sigma_\mu^2\rho}$ (189.45) and for spatial error dependence and spatial lag $RS_{\lambda\tau}$ (30.57) are highly significant after comparing them to χ_2^2 critical points. These joint tests are however not informative about the specific direction(s) of the misspecification(s). All the unadjusted statistics $RS_{\sigma_\mu^2}, RS_\rho, RS_\tau$ and RS_λ strongly reject the respective null hypothesis. If an investigator takes these rejections at their face values, then s/he would attempt to incorporate these four parameters into the final model. However, as we pointed out these one-directional tests are not valid in presence of other possible effects. Significance of each parameter can only be evaluated correctly by considering our modified tests. Three of the modified versions $RS^*_{\sigma_\mu^2}$, RS^*_ρ and RS^*_λ still reject the respective null at 1 % significance level, when compared to χ^2_2 critical values, though it is interesting to see how the values of the statistics reduce after modification. A somewhat striking result, that the value of RS^*_τ is 0.076 in contrast to that of RS_τ which is 5.0332. From our analytical results in the previous section it is clear that 5.0332 is not only for the spatial lag dependence but also reflect the presence of error dependence which seems to be much stronger for this data set. Thus the misspecification of the basic model can be thought to come from the presence of random effects, serial correlation and spatial error (rather than spatial lag dependence) of the real income of the countries.

This example seems to illustrate clearly the main points of the paper: the proposed modified versions of RS tests are more informative than the unmodified counterparts. It is worth noting a few observations from our analytical and the empirical results. Since $RS_{\sigma_\mu^2\rho\lambda\tau} = RS_{\sigma_\mu^2\rho} + RS_{\lambda\tau}$, joint test for serial correlation and random effect is independent of the joint test for spatial lag and spatial error dependence. However, further additivity fails, as we note: $RS_{\sigma_\mu^2\rho} \neq RS_{\sigma_\mu^2} + RS_\rho$ and $RS_{\lambda\tau} \neq RS_\lambda + RS_\tau$. This is due to the non-zero interaction effects between parameters, and thus unadjusted statistics are contaminated by the presence of other parameters.

From Tables 3.1 - 3.2, we have the following:

$$RS_\rho + RS_{\sigma_\mu^2} - RS_{\sigma_\mu^2\rho} = RS_{\sigma_\mu^2} - RS^*_{\sigma_\mu^2} = RS_\rho - RS^*_\rho = 25.95$$

$$RS_\lambda + RS_\tau - RS_{\lambda\tau} = RS_\lambda - RS^*_\lambda = RS_\tau - RS^*_\tau = 25.91$$

Thus, 25.95 can be viewed as a measure of the interaction between σ_μ^2 and ρ, and similarly 25.91 is the measure for λ and τ. Also these interaction effects are equal to the correction needed for the respective unadjusted tests.

It is important to emphasize again that the implementation of the modified tests is based solely on OLS residuals and parameter estimates. Some currently available test strategies relies on maximum likelihood estimation of the general spatial panel model with all the parameters, and then carrying out likelihood ratio (LR) or conditional RS tests individually or jointly. However, we propose asymptotically equivalent tests without estimating the complex model at all. In the next section we demonstrate that though our suggested tests are valid only for large samples and local misspecification, they perform quite well in finite samples and also for not-so-local departures. We also show that a very little is lost in terms of size and power in using our simple tests compared to the full-fledge computationally demanding tests.

3.7 Monte Carlo Results

To facilitate comparisons with existing results we follow a structure close to Baltagi, Song, Jung and Koh (2007) and Baltagi and Liu (2008). The data were generated using the model:

$$y_{it} = \alpha + \tau \sum_{j=1}^{N} w_{ij} y_{jt} + X_{it}\beta + u_{it} \tag{3.38}$$

$$u_{it} = \mu_i + \epsilon_{it} \tag{3.39}$$

$$\epsilon_{it} = \lambda \sum_{j=1}^{N} w_{ij}\epsilon_{jt} + v_{it} \tag{3.40}$$

$$v_{it} = \rho v_{it-1} + e_{it}, \text{where } e_{it} \sim IIDN(0, \sigma_e^2) \tag{3.41}$$

We set $\alpha = 5$ and $\beta = 0.5$. The independent variable X_{it} is generated using:

$$X_{it} = 0.4X_{it-1} + \varphi_{it} \tag{3.42}$$

where $\varphi \sim Unif[-0.5, 0.5]$ and $X_{i0} = 5 + 10\varphi_{i0}$. For weight matrix W, I consider rook design. I fixed $\sigma_\mu^2 + \sigma_e^2 = 20$ and let $\eta = \frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_e^2}$. Values of all the four parameters σ_μ^2, ρ, τ and λ are varied over a range from 0 to 0.5 . I have considered two different pairs for (N, T) namely $(25, 12), (49, 12), (25, 7), (49, 7)$. For lack of space I report the results for

$(25, 12), (49, 12)$. The results for other pairs are quite comparable to the reported ones and are available on request. Each Monte Carlo experiment is consist of generating 1000 samples for each different parameter settings. Thus the maximum standard error of the estimates of the size and power would be $\sqrt{\frac{(0.5(1-0.5)}{1000}} = 0.015$. The parameters were estimated using OLS, and eleven test statistics, namely $RS_{\sigma^2_\mu\rho\lambda\tau}, RS_{\sigma^2_\mu\rho}, RS_{\tau\lambda}, RS^*_{\sigma^2_\mu}, RS_{\sigma^2_\mu}, RS^*_\rho, RS_\rho, RS^*_\tau$, RS_τ, RS^*_λ and RS_λ were computed. As discussed earlier, in practice we do not need to compute all these statistics; we do it here for comparative evaluation. The tables and graphs are based on the nominal size of 0.05. In order to elaborate our results systematically, we divided the results in two sections. In Section 2.6.1, we present the Monte Carlo results for $RS_{\sigma^2_\mu\rho\lambda\tau}, RS_{\sigma^2_\mu\rho}, RS_{\tau\lambda}, RS^*_{\sigma^2_\mu}, RS_{\sigma^2_\mu}, RS^*_\rho$ and RS_ρ, i.e., the different test statistics for the autocorrelation and individual random effects, both in presence and absence of spatial parameters λ and τ, and in Section 3.6.2, the rest of the results are reported.

3.7.1 Monte Carlo Results for Tests relating to σ^2_μ and ρ

We discuss the results for the following parameter settings: i) $\lambda = 0$ and $\tau = 0$, i.e., when there is no spatial dependence; a case similar to Baltagi and Li (1995) and Bera et.al (2001). Our results are reported in Tables 3.3 and 3.4.

In Table 3.5 and Table 3.6 we reproduce some of the given results for the conditional LM test, from Baltagi et al. (2007, pp. 8-9), i.e., for one-dimensional conditional test for $H_0 : \rho$ in presence of σ^2_μ and λ , which are very similar when one have $\lambda = 0$ and $\lambda \neq 0$. This supports our mathematical results further that the inference on ρ is affected only by the presence of random effect, not by the presence of spatial dependence. We had provided the analytical justification for that in Section 2.3. From Table 3.5 we note that our RS^*_ρ and Baltagi et al. $RS_{\rho|\sigma^2_\mu\lambda}$ have similar performance, though the former does not require any estimation of σ^2_μ and λ. For example, when $\rho = 0$ and $\eta = 0$, then the rejection probability of RS^*_ρ is 0.054 and that of $RS_{\rho|\sigma^2_\mu\lambda}$ is 0.062. Even when η is 0.5 then RS^*_ρ is 0.055 and $RS_{\rho|\sigma^2_\mu\lambda}$ is 0.63. In terms of power also there is not much difference between the two.

By assuming $\lambda = 0$ and $\tau \neq 0$, and conducting the Monte Carlo experiments for our adjusted test statistics $RS^*_{\sigma^2_\mu}$ and RS^*_ρ, we can show, as in Table 2.6, that our experiment results are comparable to Montes-Rojas (2010) and Baltagi and Liu (2008).

ii) $\lambda = 0.3$ and $\tau = 0.05$ (i.e., in local presence of the spatial lag and error dependence). iii) $\lambda = 0.05$ and $\tau = 0.3$ (i.e., when there is local presence of the spatial parameters). Figures 2.1 to Figure 2.8 illustrate the size and power of our adjusted test statistics $RS^*_{\sigma^2_\mu}$ and RS^*_ρ.

Let us now consider the performances of the tests in terms of power and sizes of RS^*_ρ and $RS^*_{\sigma^2_\mu}$. For N=25, T=12, the estimated rejection probabilities are reported in Table 3.2, and for N=49, T=12 it is reported in Table 2.3. For both of these tables, the estimated rejection probabilities are for data generated with $\lambda = 0$ and $\tau = 0$. We also illustrate the results of the other two cases [(ii) and (iii)] graphically from Figures 3.1- 3.16. Let us first concentrate on $RS^*_{\sigma^2_\mu}$ and $RS_{\sigma^2_\mu}$, which are designed to test $H_0 : \sigma^2_\mu = 0$. When $\rho = 0$, there is a loss of power for $RS^*_{\sigma^2_\mu}$ vis-a-vis $RS_{\sigma^2_\mu}$, and this loss gets minimized as η deviates more and more from zero. While $RS^*_{\sigma^2_\mu}$ does not sustain much loss in power when $\rho = 0$, we notice that $RS_{\sigma^2_\mu}$ reject $H_0 : \sigma^2_\mu = 0$ too often when $\sigma^2_\mu = 0$, but $\rho \neq 0$. This unwanted rejection probabilities is due to presence of ρ, as shown in equation (3.26). $RS^*_{\sigma^2_\mu}$ also has some rejection probabilities but the problem is less severe as can be seen from equation (3.26). We showed in Table 3.6 that our results are similar as in Table 3.3 in Baltagi et al. (2007). One dimensional LM and LR tests, as in Table 3.3 in Baltagi et. al (2007), are obtained after estimating the parameters σ^2_μ and λ . However, RS^*_ρ in Table 3.6 illustrates that one can obtain very similar results, like conditional LM or LR tests, using our adjusted RS test.

Moreover, even when we allow for case ii) $\lambda = 0.3$ and $\tau = 0.05$, case iii) $\lambda = 0.05$ and $\tau = 0.3$; we find close results as in Table 3.3 and 3.4 above, i.e., $RS^*_{\sigma^2_\mu}$ are robust only under the local misspecification, i.e., for low values of ρ. From Figures 3.1 and 3.2 we can see that $RS^*_{\sigma^2_\mu}$ is size robust for local misspecification of the parameters under both the cases ii) $\lambda = 0.3$ and $\tau = 0.05$, and iii) $\lambda = 0.05$ and $\tau = 0.3$. Comparing the power (Figures 3.3 - 3.4), it is clearly evident that the power loss gets minimized for $RS^*_{\sigma^2_\mu}$ as η deviates from 0.

The Figures 3.1 - 3.4 confirm that the local presence of the spatial dimensions do not affect $RS^*_{\sigma^2_\mu}$ drastically, which confirms our mathematical proof. The features of $RS^*_{\sigma^2_\mu}$ is more or less similar when $\lambda = 0$ and $\tau = 0$ vis-a-vis their local departures from 0. It means that the inference of the parameter σ^2_μ does not depend on the local presence of spatial parameters λ and τ.

In similar way, we can explain the size and power of RS_ρ^* using Table 3.3 and 3.4 and Figures 3.5 - 3.8. From Table 3.3 and 3.4 we note that when $\sigma_{\mu^2} = 0$ then RS_ρ has better power than RS_ρ^* . However, unlike $RS_{\sigma_\mu^2}^*$, the power of RS_ρ^* is much closer to RS_ρ. The real benefit of RS_ρ^* is noticed when $\rho = 0$ but $\eta > 0$; the performance of RS_ρ^* is remarkable. For the case $\lambda = 0, \tau = 0$, for N=25, T=12 and N=49, T=12 it is evident from the Tables 3.3 and Table 3.4. Even when there is local presence of the parameters λ and τ, the size of RS_ρ^* is significantly better than RS_ρ when $\eta > 0$. In other words, RS_ρ^* is performing much better than what it is expected to perform; i.e. not rejecting $\rho = 0$ when ρ is indeed zero even for large values of η. The non central parameter of RS_ρ^* , as shown in equation (3.28) is independent of any nuisance parameters, which explains the robust performance of the test statistics. On the other hand RS_ρ rejects the null too often even when ρ is actually zero. This can be easily observed from equation (3.28) which is a function of σ_μ^2. The power comparison also gives a very impressive result.

Results from the joint statistics $RS_{\sigma_\mu^2\rho\lambda\tau}$ and $RS_{\sigma_\mu^2\rho}$ is informative when we accept the respective null hypotheses $H_0 : \sigma_\mu^2 = \rho = \lambda = \tau = 0$ and $H_0 : \sigma_\mu^2 = \rho = 0$ respectively. However, if the null is rejected we need to decompose $RS_{\sigma_\mu^2\rho\lambda\tau}$ and $RS_{\sigma_\mu^2\rho}$ to extract exact source(s) of misspecification. However, overall they have good power. These results are consistent with Bera et al. (2001) and also Montes-Rojas (2010). However, we differ from each of them in our basic model framework, which is more general than both Bera et al. (2001) and Montes-Rojas (2010).

Table 3.5 illustrates that the Monte Carlo results for the conditional LM test, derived in Baltagi et al. (2007, p. 8-9), i.e., one dimensional conditional test for C.2, $H_0^i : \rho = 0$ in presence of σ_μ^2 and λ, which are very similar when $\lambda = 0$ and $\lambda \neq 0$. This supports our mathematical results further: the inference on ρ is affected only by the presence of random effect, not by the presence of spatial dependence. Infact the results in Table 3.5 is close to the Monte Carlo results in Baltagi and Li (1995) which illustrated one dimensional conditional test of ρ in presence of σ_μ^2 only. Table 3.2 in Baltagi and Li (1995) report the power of $RS_{\rho|\sigma_\mu^2}$ after estimating σ_μ^2. For example, the Monte Carlo results for $\rho = 0.4$ and $\sigma_\mu^2 = 0$ is 1.000, which is comparable to $RS_{\rho|\sigma_\mu^2\lambda}$ of Baltagi et al. (2007) where $RS_{\rho|\sigma_\mu^2\lambda}$ is calculated after estimating σ_μ^2 and λ .This illustrates our findings further that asymptotically the LM or RS test statistics of time series parameters, i.e. ρ and σ_μ^2 are independent of spatial parameters i.e. λ and τ . Even in the finite sample the size and power do not differ much as illustrated in our Table 3.5. We also conducted the Monte Carlo experiments for our adjusted RS test for error autocorrelation, i.e., RS_ρ^*, assuming $\lambda \neq 0$ and $\tau = 0$ and compared it with the one dimensional LM test derived in Baltagi et al. (2007).

Here $RS_{\rho|\sigma_\mu^2\lambda}$ refers to the one dimensional conditional LM test as derived in Baltagi et al. (2007). The rejection probabilities for $RS_{\rho|\sigma_\mu^2\lambda}$ are the ones reported in Baltagi et. al (2007) in Table 3.3 for N=25, T=12. We computed RS_ρ^* for $\lambda \neq 0$ and $\eta > 0$. As noted before RS_ρ^* can be computed using simple OLS residuals, whereas computation of $RS_{\rho|\sigma_\mu^2\lambda}$ requires estimation of λ and σ_μ^2. Results reported in Table 3.6 further supports our findings, i.e., on one hand the performance of our adjusted RS is very similar to one directional conditional LM test; on the other, our adjusted RS test is simple to compute than conditional LM test.

3.7.2 Monte Carlo Results for Tests relating to τ and λ

Let us now consider the parameters of spatial dimensions. To explore the performance of these tests we have performed the Monte Carlo study for three cases: i) $\eta = \rho = 0$ (This case is exactly similar to Anselin et. al. (1996), and our results are comparable to their findings.) ii) $\eta = 0.05$ and $\rho = 0.3$ iii) $\eta = 0.3$ and $\rho = 0.05$ The results of the last two cases are comparable to the Monte Carlo results of Baltagi et al. (2007) and Baltagi and Liu (2008). Table 3.7 and Table 3.8 give the estimated rejection probabilities of the tests $RS_\lambda^*, RS_\lambda, RS_\tau^*, RS_\tau, RS_{\sigma_\mu^2\rho\lambda\tau}$, and $RS_{\lambda\tau}$, for the case $\eta = \rho = 0$ for sample size (N,T):(25,12),(49,12) respectively. The RS_λ has power against a 'spatial lag', although less than the lag tests i.e., RS_τ. The behavior of RS_λ^* is interesting. It has no power against lag dependence i.e. τ, as it should. For small values of τ, the rejection frequency of RS_λ^* is very close to its expected value of 0.05. In fact for (N, T =49,12) this size robustness of RS_λ^* is more evident as the rejection frequency is close to 0.05 even when$\tau = 0.5$. In other words, RS_λ^* does its job very well, even more than what it is designed to do for. However the rejection frequency of RS_λ is large in presence of τ even when λ is actually equal to zero. This reiterates our result that RS_λ^* is robust to 'local' misspecification, while the test results of RS_λ can be very misleading in presence of such nuisance parameters. In terms of power, RS_λ^* is trailing just behind RS_λ as can be clearly seen from the tables. From Table 3.7 and Table 3.8, it is evident that RS_τ^* is size robust for local misspecification of λ. For $\lambda > 0$ and $\tau = 0$, RS_τ has rejection probabilities higher than 0.05, but it is much less than RS_τ^*. This unwanted rejection probabilities of RS_τ is due to the non-centrality term which depends on λ. As mentioned before, RS_τ^* is designed to be robust only under local misspecification, i.e., for low values of λ. From that point of view, it does a good job; the performances deteriorate as λ takes higher values. From the tables we also note that when $\tau > 0$ an increase in λ enhances the rejection probabilities of RS_τ . This is due to the presence of λ in the non-centrality parameter of RS_τ. But the non-centrality parameter of RS_τ^* does not depend on λ (Proofs regarding non centrality parameters of these tests can be found in Bera and Yoon (1993)).This result is valid only asymptotically and for local departures of λ.

We further investigated the behavior of RS_λ^* for the two other cases, i.e., in local presence of the error autocorrelation and random effect: ii) $\eta = 0.05$ and $\rho = 0.3$ iii) $\eta = 0.3$ and $\rho = 0.05$ The results are explained through the Figures 3.9 to 3.16. The size of RS_λ^* is much better than its unadjusted counterpart in local presence of all the three parameters η, ρ and τ. The power of RS_λ^* is slightly less than that of RS_λ, given $\tau = 0$ for both the above cases. The Figures 3.9 - 3.12, clearly show that the rejection probabilities are very close to 0.05 for RS_λ^* for τ varying from 0, 0.05, 0.1, 0.3 and 0.5. The rejection probability of RS_λ^* increases with λ as it should be and can be explained by equation (3.30). On the other hand the rejection probability of RS_λ is always very high even when $\lambda = 0$ and τ being away from zero. This can be explained using the non centrality parameter in presence of nuisance parameters as shown in equation (3.30). This re-iterates our earlier result as evident from Table 3.5 and 3.6, when both the random region effect and error autocorrelation effect were absent i.e, $\eta = \rho = 0$. These experimental results provide further support to our mathematical findings.

Finally, we discuss the experimental results of RS_τ^* and RS_τ in local presence of ρ and η, i.e., for the following two cases: ii) $\eta = 0.05$ and $\rho = 0.3$ iii) $\eta = 0.3$ and $\rho = 0.05$ Figures 3.13 - 3.16 explains the results of the tests for the above two cases. The size of the test RS_τ^* is robust for the local presence of λ; and it increases to approximately 0.4 when $\lambda = 0.5$. In contrast the size of RS_τ is approaches 1 when $\lambda = 0.5$. These rejection probabilities can be explained by equation (2.32). Thus although there is some unwanted rejection probability problem with RS_τ^* (as discussed before) still the problem is much less severe of RS_τ^* than RS_τ; even when there is local presence of η and ρ. The power of RS_τ^* trails behind RS_τ, but becomes c lose to each other for larger values of τ. These results is in similar lines of what we found when η and ρ were zero, i.e., the local presence of the random effect and time series error autocorrelation do not influence the inference of these tests. Thus these once again support our analytical finding regarding these tests.

One important thing to note is these one dimensional size-robust tests are more meaningful not only from their marginal counterparts but also from the joint tests, $RS_{\sigma_\mu^2\rho\lambda\tau}$ (four dimensional), $RS_{\sigma_\mu^2\rho}$, and $RS_{\lambda\tau}$ (two dimensional) tests. These joint tests are only optimal when $\sigma_\mu^2 = \rho = \lambda = \tau = 0$. These tests fail to identify the exact source of misspecification. This is evident from Tables 3.3 - 3.4 and 3.7 - 3.8. In addition, as stated before our conditional tests not only give intuitive results which one can explain analytically and also mathematically, but these are also easy to compute (they are all based on OLS estimates) than the one dimensional conditional LM and LR tests. Moreover, one can easily derive our adjusted test statistic using the unadjusted ones (Eq. 3.29 - 3.33) as shown in Section 3.4.

3.8 Conclusion

Based only on OLS estimation, in this paper we have proposed "robust" Rao's Score (RS) test for random effect, serial correlation, spatial error and spatial lag dependence in the context of spatial panel model. The tests are robust in the sense that they are asymptotically valid in the (local) presence of nuisance parameters. After one has the standard RS tests for each parameter, our robust tests require very little extra computation. Thus practitioners can identify specific direction(s) to reformulate the basic model without going through any complex estimation. Our empirical illustration in the context of the convergence theory of income of different countries demonstrates the usefulness of our proposed tests, in particular, to identify the exact form of spatial dependence. We also have investigated the finite sample size and a power property of our proposed tests through an extensive Monte Carlo study, and compared them with the performance of some of the available tests. Our tests perform very well in finite sample and compares favorably to other tests that require explicit estimation of nuisance parameters. Also, though our methodology is developed only for local misspecification, results from our simulation experiments show that in certain cases, our tests perform quite well for nonlocal departures.

3.9 Tables and Figures

Table 3.1: Joint Tests

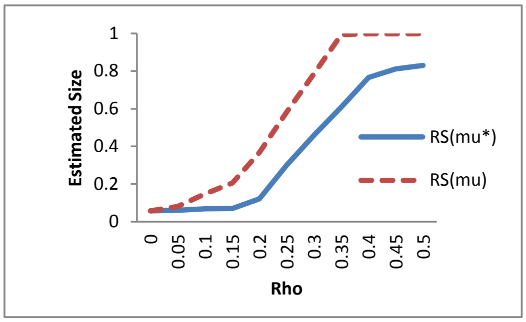
$RS_{\sigma_\mu^2\rho\lambda\tau}$	$RS_{\sigma_\mu^2\rho}$	$RS_{\lambda\tau}$
220.02	189.45	30.57

Table 3.2: Tests for random effect, serial correlation, spatial error and spatial lag

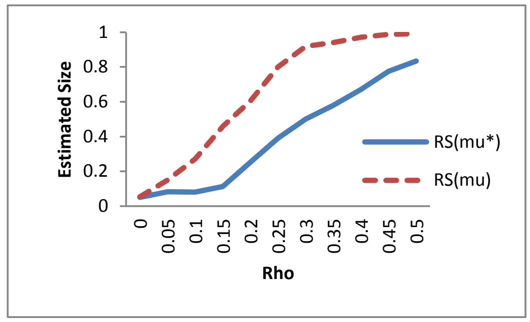
$RS_{\sigma_\mu^2}$	$RS^*_{\sigma_\mu^2}$	RS_ρ	RS^*_ρ	RS_λ	RS^*_λ	RS_τ	RS^*_τ
183.03	157.14	32.31	6.36	26.01	0.10	30.46	4.55

Size Comparison of RS (mu) and RS (mu*)

3.1) $\lambda = 0.05,\ \tau = 0.3$ and $\sigma_\mu{}^2 = 0$

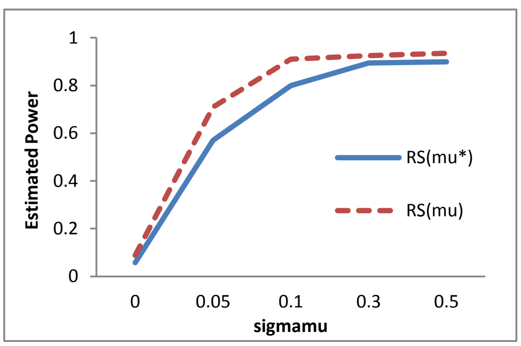


3.2) $\lambda = 0.3,\ \tau = 0.05$ and $\sigma_\mu{}^2 = 0$

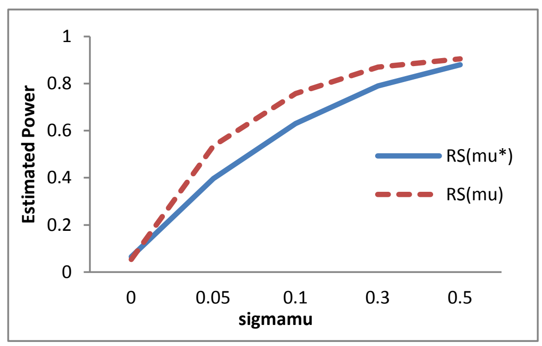


Power Comparison

3.3) $\lambda = 0.05$ and $\tau = 0.3$

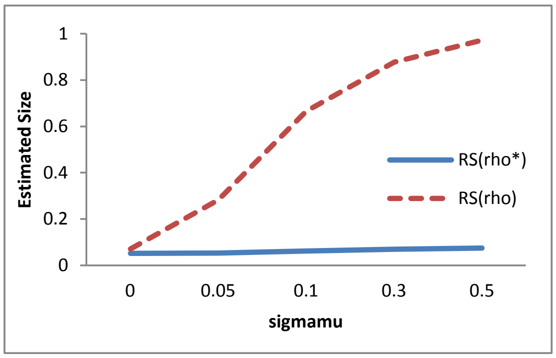


3.4) $\lambda = 0.3$ and $\tau = 0.05$

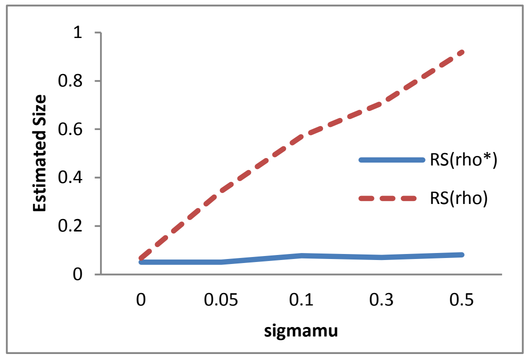


Size Comparison of RS (rho) and RS (rho*)

3.5) $\lambda = 0.05$, $\tau = 0.3$ and $\rho = 0$

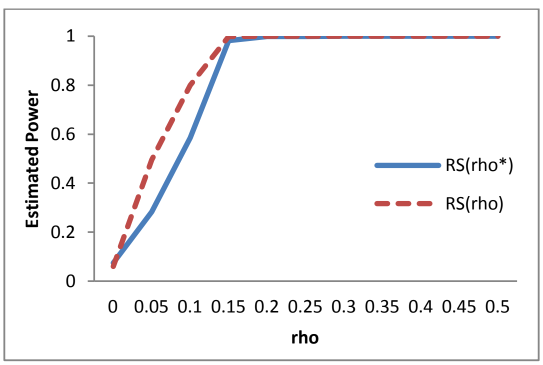


3.6) $\lambda = 0.3$, $\tau = 0.05$ and $\rho = 0$



Power Comparison

3.7) $\lambda = 0.05$ and $\tau = 0.3$



3.8) $\lambda = 0.3$ and $\tau = 0.05$

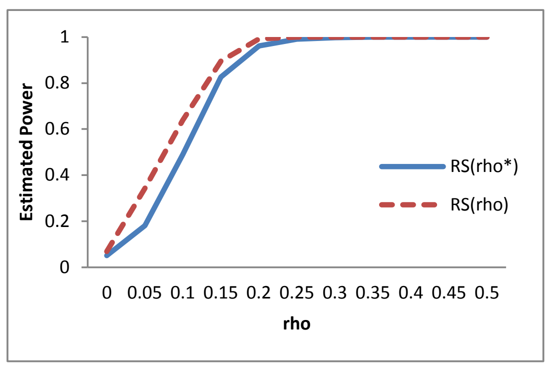


Table 3.3: Estimated Rejection Probabilities with $\tau = \lambda = 0$. Sample size: $N = 25, T = 10$

η	ρ	$RS_{\sigma_\mu^2 \rho \lambda \tau}$	$RS_{\sigma_\mu^2 \rho}$	$RS_{\sigma_\mu^2}$	$RS^*_{\sigma_\mu^2}$	RS_ρ	RS^*_ρ
0	0	0.047	0.069	0.054	0.048	0.050	0.049
0	0.05	0.609	0.858	0.068	0.057	0.188	0.122
0	0.1	0.818	0.943	0.130	0.082	0.670	0.652
0	0.2	0.985	0.999	0.580	0.390	0.999	0.985
0	0.3	1.000	1.000	0.938	0.482	1.000	1.000
0	0.4	1.000	1.000	0.974	0.614	1.000	1.000
0	0.5	1.000	1.000	0.999	0.781	1.000	1.000
0.05	0	0.399	0.681	0.547	0.209	0.281	0.059
0.05	0.05	0.621	0.839	0.614	0.293	0.577	0.320
0.05	0.1	0.786	0.942	0.682	0.305	0.858	0.505
0.05	0.2	0.974	0.998	0.836	0.380	1.000	0.983
0.05	0.3	1.000	1.000	0.925	0.463	1.000	0.998
0.05	0.4	1.000	1.000	0.972	0.611	1.000	1.000
0.05	0.5	1.000	1.000	0.993	0.781	1.000	1.000
0.1	0	0.381	0.686	0.543	0.400	0.687	0.057
0.1	0.05	0.582	0.838	0.605	0.412	0.854	0.592
0.1	0.1	0.800	0.944	0.706	0.600	0.961	0.800
0.1	0.2	0.984	0.996	0.829	0.682	0.999	0.977
0.1	0.3	0.998	1.000	0.930	0.761	1.000	0.997
0.1	0.4	1.000	1.000	0.969	0.84	1.000	1.000
0.1	0.5	1.000	1.000	0.994	0.98	1.000	1.000
0.3	0	0.440	0.704	0.789	0.705	0.701	0.051
0.3	0.05	0.572	0.822	0.856	0.811	0.832	0.552
0.3	0.1	0.719	0.903	0.985	0.844	0.926	0.800
0.3	0.2	0.913	0.978	1.000	0.973	0.990	0.921
0.3	0.3	0.996	1.000	1.000	1.000	1.000	0.989
0.3	0.4	1.000	1.000	1.000	1.000	1.000	0.994
0.3	0.5	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0	0.431	0.719	1.000	0.905	0.709	0.053
0.5	0.05	0.536	0.796	1.000	0.911	0.814	0.182
0.5	0.1	0.648	0.859	1.000	0.944	0.878	0.432
0.5	0.2	0.841	0.961	1.000	0.973	0.970	0.835
0.5	0.3	0.963	0.991	1.000	1.000	0.997	0.946
0.5	0.4	0.995	1.000	1.000	1.000	0.999	0.982
0.5	0.5	1.000	1.000	1.000	1.000	1.000	0.989

Table 3.4: Estimated Rejection Probabilities with $\tau = \lambda = 0$. Sample size: $N = 49, T = 12$

η	ρ	$RS_{\sigma_\mu^2 \rho\lambda\tau}$	$RS_{\sigma_\mu^2 \rho}$	$RS_{\sigma_\mu^2}$	$RS^*_{\sigma_\mu^2}$	RS_ρ	RS^*_ρ
0	0	0.060	0.059	0.053	0.047	0.073	0.051
0	0.05	0.935	0.991	0.415	0.054	0.992	0.797
0	0.1	1.000	0.999	0.853	0.056	1.000	0.992
0	0.2	1.000	1.000	0.986	0.187	1.000	1.000
0	0.3	1.000	1.000	0.996	0.725	1.000	1.000
0	0.4	1.000	1.000	1.000	0.856	1.000	1.000
0	0.5	1.000	1.000	1.000	0.949	1.000	1.000
0.05	0	1.000	0.945	0.848	0.350	0.531	0.047
0.05	0.05	1.000	0.989	0.889	0.420	0.989	0.776
0.05	0.1	1.000	0.999	0.943	0.508	1.000	0.987
0.05	0.2	1.000	1.000	0.990	0.617	1.000	1.000
0.05	0.3	1.000	1.000	0.998	0.696	1.000	1.000
0.05	0.4	1.000	1.000	1.000	0.847	1.000	1.000
0.05	0.5	1.000	1.000	1.000	0.943	1.000	1.000
0.1	0	1.000	0.928	0.850	0.519	0.924	0.051
0.1	0.05	1.000	0.988	0.894	0.533	0.988	0.884
0.1	0.1	1.000	0.999	0.941	0.537	1.000	0.977
0.1	0.2	1.000	1.000	0.987	0.579	1.000	1.000
0.1	0.3	1.000	1.000	0.995	0.713	1.000	1.000
0.1	0.4	1.000	1.000	1.000	0.838	1.000	1.000
0.1	0.5	1.000	1.000	1.000	0.943	1.000	1.000
0.3	0	1.000	0.953	0.855	0.760	0.938	0.054
0.3	0.05	1.000	0.984	0.904	0.812	0.983	0.871
0.3	0.1	1.000	0.994	0.930	0.942	0.997	0.949
0.3	0.2	1.000	1.000	1.000	0.990	1.000	0.999
0.3	0.3	1.000	1.000	1.000	1.000	1.000	1.000
0.3	0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.3	0.5	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0	1.000	0.961	1.000	1.000	0.948	0.045
0.5	0.05	1.000	0.982	1.000	1.000	0.980	0.810
0.5	0.1	1.000	0.989	1.000	1.000	0.996	0.918
0.5	0.2	1.000	1.000	1.000	1.000	1.000	0.987
0.5	0.3	1.000	1.000	1.000	1.000	1.000	0.998
0.5	0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.5	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.5: Estimated rejection probabilities of RS^*_ρ, $RS_{\rho|\sigma^2_\mu\lambda}$ with $\lambda = 0$. Sample Size: N=25,T=12

	$\eta = 0$		$\eta = 0.2$		$\eta = 0.5$	
ρ	RS^*_ρ	$RS_{\rho\|\sigma^2_\mu\lambda}$	RS^*_ρ	$RS_{\rho\|\sigma^2_\mu\lambda}$	RS^*_ρ	$RS_{\rho\|\sigma^2_\mu\lambda}$
0	0.054	0.062	0.055	0.051	0.055	0.063
0.2	0.730	0.815	0.790	0.816	0.833	0.848
0.4	1.000	1.000	1.000	0.990	0.990	0.982
0.6	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000	1.000	1.000
0.9	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.6: Estimated rejection probabilities of RS^*_ρ, $RS_{\rho|\sigma^2_\mu\lambda}$ with $\lambda \neq 0$. Sample Size: N=25,T=12

		$\eta = 0$		$\eta = 0.2$		$\eta = 0.5$	
λ	ρ	RS^*_ρ	$RS_{\rho\|\sigma^2_\mu\lambda}$	RS^*_ρ	$RS_{\rho\|\sigma^2_\mu\lambda}$	RS^*_ρ	$RS_{\rho\|\sigma^2_\mu\lambda}$
0	0	0.054	0.051	0.053	0.053	0.055	0.070
0.2	0	0.056	0.061	0.051	0.061	0.051	0.045
0.4	0	0.058	0.047	0.055	0.053	0.047	0.035
0.6	0	0.059	0.051	0.057	0.056	0.037	0.042
0.8	0	0.055	0.041	0.046	0.046	0.031	0.031
0	0.	0.730	0.803	0.790	0.817	0.833	0.848
0.2	0.2	0.790	0.785	0.810	0.827	0.814	0.817
0.4	0.2	0.865	0.842	0.802	0.812	0.828	0.810
0.6	0.2	0.802	0.826	0.816	0.816	0.802	0.810
0.8	0.2	0.775	0.813	0.755	0.814	0.745	0.810
0	0.4	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.7: Estimated Rejection Probabilities with $\eta = \rho = 0$. Sample size: $N = 25, T = 12$

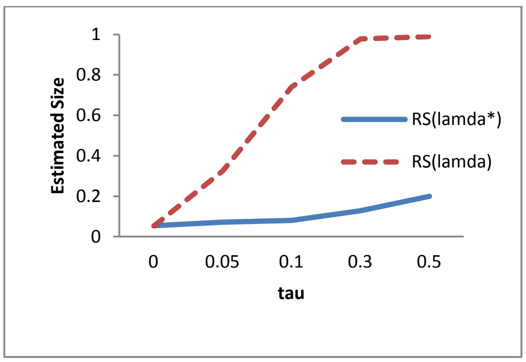
λ	τ	$RS_{\sigma_\mu^2\rho\lambda\tau}$	$RS_{\lambda\tau}$	RS_λ	RS_λ^*	RS_τ	RS_τ^*
0	0	0.059	0.055	0.051	0.056	0.051	0.052
0.05	0	0.797	0.397	0.231	0.184	0.774	0.054
0.1	0	0.998	0.543	0.378	0.288	0.896	0.060
0.2	0	1.000	0.998	0.596	0.329	0.999	0.141
0.3	0	1.000	1.000	0.865	0.587	1.000	0.159
0.4	0	1.000	1.000	0.910	0.745	1.000	0.164
0.5	0	1.000	1.000	0.996	0.889	1.000	0.191
0	0.05	0.099	0.158	0.258	0.059	0.386	0.151
0.05	0.05	0.999	0.327	0.482	0.373	0.595	0.140
0.1	0.05	1.000	0.523	0.695	0.567	0.799	0.117
0.2	0.05	1.000	0.867	1.000	0.967	0.995	0.205
0.3	0.05	1.000	0.992	1.000	0.996	1.000	0.228
0.4	0.05	1.000	1.000	1.000	0.999	1.000	0.273
0.5	0.05	1.000	1.000	1.000	1.000	1.000	0.297
0	0.1	1.000	0.069	0.683	0.055	0.790	0.360
0.05	0.1	1.000	0.356	0.893	0.275	0.899	0.461
0.1	0.1	1.000	0.603	0.999	0.309	0.900	0.596
0.2	0.1	1.000	0.834	1.000	0.876	1.000	0.620
0.3	0.1	1.000	0.933	1.000	0.998	1.000	0.635
0.4	0.1	1.000	1.000	1.000	1.000	1.000	0.691
0.5	0.1	1.000	1.000	1.000	1.000	1.000	0.696
0	0.3	1.000	0.180	0.399	0.082	0.856	0.666
0.05	0.3	1.000	0.390	0.487	0.386	0.980	0.747
0.1	0.3	1.000	0.650	0.698	0.597	0.999	0.802
0.2	0.3	1.000	0.878	1.000	0.999	1.000	0.907
0.3	0.3	1.000	1.000	1.000	1.000	1.000	0.966
0.4	0.3	1.000	1.000	1.000	1.000	1.000	0.987
0.5	0.3	1.000	1.000	1.000	1.000	1.000	0.999
0	0.5	1.000	1.000	0.432	0.048	1.000	0.929
0.05	0.5	1.000	1.000	0.790	0.543	1.000	0.943
0.1	0.5	1.000	1.000	0.876	0.699	1.000	0.950
0.2	0.5	1.000	1.000	0.998	0.878	1.000	0.999
0.3	0.5	1.000	1.000	1.000	1.000	1.000	1.000
0.4	0.5	1.000	1.000	1.000	1.000	1.000	0.997
0.5	0.5	1.000	1.000	1.000	1.000	1.000	0.996

Table 3.8: Estimated Rejection Probabilities with $\eta = \rho = 0$. Sample size: $N = 49, T = 12$

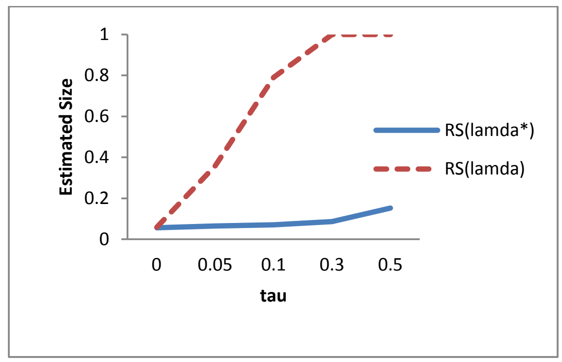
λ	τ	$RS_{\sigma_\mu^2 \rho\lambda\tau}$	$RS_{\lambda\tau}$	RS_λ	RS_λ^*	RS_τ	RS_τ^*
0	0	0.052	0.055	0.052	0.076	0.051	0.062
0.05	0	1.000	0.122	0.261	0.396	0.053	0.374
0.1	0	1.000	0.365	0.597	0.678	0.056	0.596
0.2	0	1.000	0.567	0.789	0.823	0.087	0.999
0.3	0	1.000	0.990	0.866	0.935	0.131	1.000
0.4	0	1.000	1.000	0.976	0.999	0.159	1.000
0.5	0	1.000	1.000	0.999	1.000	0.173	1.000
0	0.05	1.000	0.076	0.053	0.279	0.150	0.350
0.05	0.05	1.000	0.187	0.384	0.594	0.178	0.560
0.1	0.05	1.000	0.432	0.557	0.689	0.284	0.739
0.2	0.05	1.000	0.765	0.896	0.967	0.308	0.995
0.3	0.05	1.000	0.990	0.956	0.999	0.369	1.000
0.4	0.05	1.000	1.000	0.990	1.000	0.496	1.000
0.5	0.05	1.000	1.000	1.000	1.000	0.598	1.000
0	0.1	1.000	0.063	0.050	0.367	0.585	0.750
0.05	0.1	1.000	0.246	0.487	0.698	0.663	0.846
0.1	0.1	1.000	0.534	0.569	0.798	0.763	0.997
0.2	0.1	1.000	0.898	0.789	0.899	0.920	1.000
0.3	0.1	1.000	0.970	0.887	0.989	0.986	1.000
0.4	0.1	1.000	1.000	0.980	1.000	0.999	1.000
0.5	0.1	1.000	1.000	1.000	1.000	1.000	1.000
0	0.3	1.000	0.324	0.049	0.494	0.773	0.998
0.05	0.3	1.000	0.521	0.542	0.799	0.835	1.000
0.1	0.3	1.000	0.876	0.715	0.870	0.890	1.000
0.2	0.3	1.000	1.000	0.987	0.999	0.968	1.000
0.3	0.3	1.000	1.000	1.000	1.000	0.994	1.000
0.4	0.3	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.3	1.000	1.000	1.000	1.000	1.000	1.000
0	0.5	1.000	1.000	0.045	0.546	0.834	1.000
0.05	0.5	1.000	1.000	0.532	0.787	0.873	1.000
0.1	0.5	1.000	1.000	0.723	0.886	0.919	1.000
0.2	0.5	1.000	1.000	0.998	1.000	0.962	1.000
0.3	0.5	1.000	1.000	1.000	1.000	0.984	1.000
0.4	0.5	1.000	1.000	1.000	1.000	0.998	1.000
0.5	0.5	1.000	1.000	1.000	1.000	1.000	1.000

Size Comparison of RS (lamda) and RS(lamda*)

3.9) $\eta = 0.05, \ \rho = 0.3$ and $\lambda = 0$

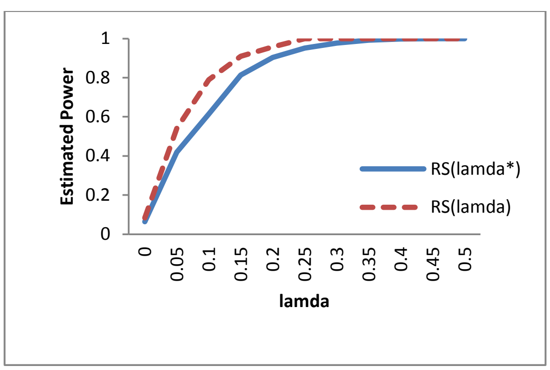


3.10) $\eta = 0.3, \ \rho = 0.05$ and $\lambda = 0$

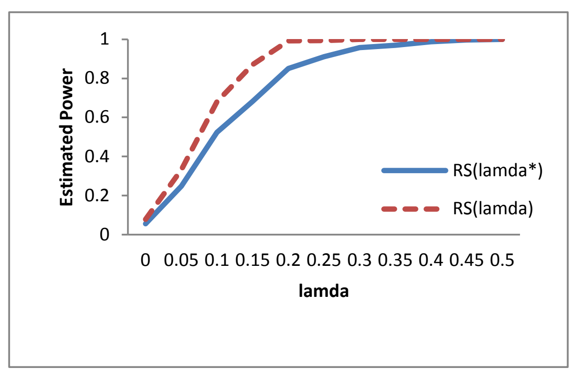


Power Comparison

3.11) $\eta = 0.05$ and $\rho = 0.3$

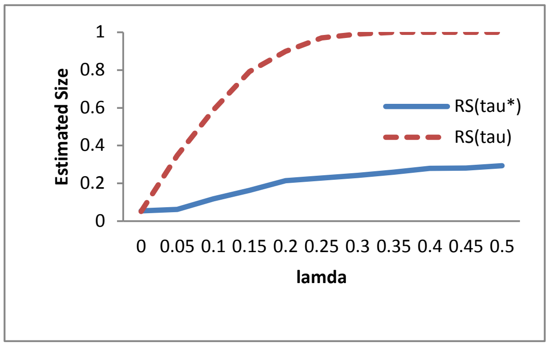


3.12) $\eta = 0.3$ and $\rho = 0.05$

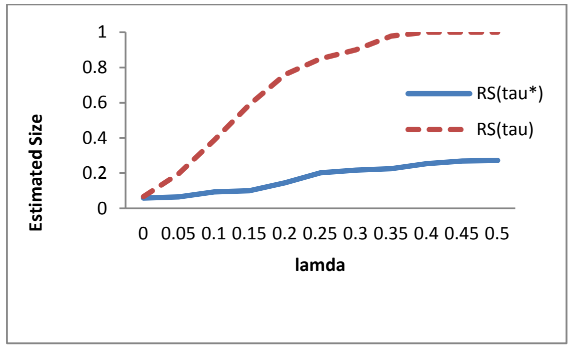


Size Comparison of RS (tau) and RS (tau*)

3.13) $\eta = 0.05,\ \rho = 0.3$ and $\tau = 0$

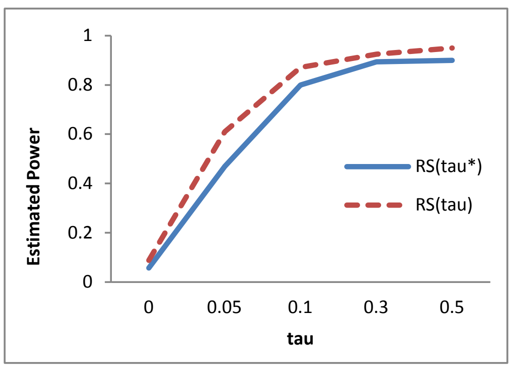


3.14) $\eta = 0.3,\ \rho = 0.05$ and $\tau = 0$

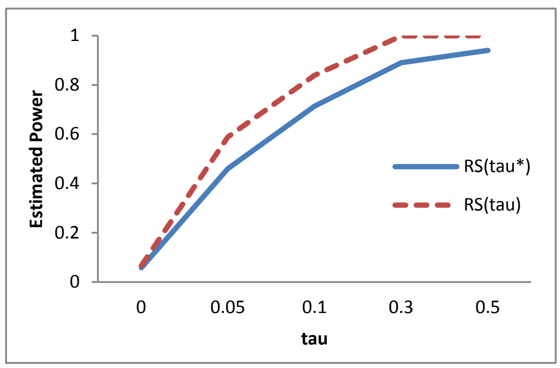


Power Comparison RS (tau) and RS (tau*)

3.15) $\eta = 0.05$ and $\rho = 0.3$



3.16) $\eta = 0.3$ and $\rho = 0.05$



CHAPTER 4

SPECIFICATION SEARCH FOR GROWTH MODEL: A DYNAMIC SPACE - TIME FRAMEWORK

4.1 Introduction

"It isn't what we don't know that kills us. It's what we know that ain't so." –Mark Twain

Understanding the nation's growth is one of the oldest and most important research agendas in economics. At the same time, the empirical study of economic growth occupies a position that is notably uneasy. Rodriguez and Rodrik (2001) begin their skeptical critique of evidence on trade policy and growth with the above quote which they use to point out the difficulty in identifying the salient determinants of growth. Quite generally, one such difficult issues is the basic econometric specification of the growth models, i.e., absence of consensus regarding the salient features of the underlying data generating process (DGP). There is a vast literature on econometric issues that arise due to different presumptions on the structure of the DGP that appear in growth analyses. The choices of method involve significant trade-offs, which depend both on statistical considerations and on economic context. In spite of the vast literature on this particular issue of econometric specification of the models, it has always been difficult to identify the structure of DGP. In this paper, I address this point. I formulate new diagnostic tests that take into account of misspecification in multiple directions. In particular, I propose adjusted Rao's score (RS) test statistics under a dynamic panel spatial model framework, which are robust under misspecification. I use the proposed tests for *specification search in multiple directions*, without any complex estimation of the nuisance parameters. The proposed test statistics can assist a researcher to revise his/her model towards appropriate direction(s) for better understanding of the growth behavior and thereby suggest suitable policy reforms.

Research on convergence proceeded through several stages and also witnessed the use of different methodologies. However, the correspondence between the convergence definitions (like β-convergence, σ-convergence, conditional convergence, and so on) and the methodologies used are not unique. For example, cross-section, panel and time series (in part) approaches have been used to study β. These approaches have generally dealt with convergence across economies and in terms of per capita income level. The cross-sectional approach is popular to study σ-convergence, while time series methodologies are implemented to investigate convergence both within and across economies. More recently, various spatial approaches has been adopted to model the technological spillover and interdependencies of economies, both in cross-sectional and panel framework. Each of these methodologies has its own benefits and drawbacks, even though it may be used to study a particular aspect of convergence. For example, the use of panel data to study β-convergence, is likely to increase efficiency and allow richer models in the presence of parameter heterogeneity. Thus there is a trade-off between robustness and efficiency with each of the chosen methods. The scientific solution would be to base the choice of estimation method on a context-specific loss function. This is clearly a very difficult task. Thus the crucial question is which model/models do the data confirm with?

This paper provides a solution to this difficult problem in the context of growth model, i.e., unravel the DGP without any subjective preference, so that the researcher can choose a suitable model to understand the underlying growth behavior. As mentioned earlier, there has been many studies that have considered only cross-section, or time series, or panel, or spatial model methods. In this paper I consider the dynamic panel spatial model framework which is a generalization of all these piece-wise models and propose test statistics to understand which kind of departures are actually present in the underlying dataset. I start with a small model (simple panel model under joint null hypothesis) and then check whether specific departures (like time dynamics, serial correlation of errors, individual effects, different forms of spatial dependencies) from this starting model are supported or rejected by the data. Using Bera and Yoon (1993) test principle, I propose new adjusted Rao's score (RS) test statistics for each of the parameters, after taking into account the possible presence of all other forms of departures. Unlike the existing conditional tests, the proposed methodology takes care of the possible presence of all the nuisance parameters through their respective Fisher-Rao score evaluated under *joint null*, and thus requiring estimation of the simplest model. Using these proposed test statistics I also show how some existing models are potentially misspecified.

The main contributions of this paper are thus twofold: (i) development of *six* new RS test statistics robust under local misspecification, i.e., adjusted RS for time dynamics, random effects, serial correlation, space-time dynamics, spatial lag and spatial dependence parameters, where each of them is robust to the presence of all the other departures (nuisance parameters). The proposed test statistics do not require any estimation of the nuisance parameters and thus, are computationally simple and easily amenable for misspecification analysis. (ii) Using these proposed test statistics, I address the empirical question of specification search, i.e., which model/models do the underlying data for growth models confirm with? Thus, using my proposed tests I come up with a proper model for the growth analysis.

The plan of the rest of the paper is as follows. The next section provides a brief review of the existing models used for growth convergence. I provide the details of our model framework and the regularity conditions in Section 4.3. In Section 4.4, I derive the new diagnostic tests which take account of misspecification in multiple directions. After reviewing the data set in Section 4.5, I discuss how the proposed test statistics can unravel the dependencies of the underlying DGP without any complex estimation of the model itself. Thus I propose an appropriate growth model that capture the salient features of the data. In Section 4.6, I conduct finite sample study to evaluate the performance of the suggested and some available tests, and Section 4.7 concludes the paper. All the figures and tables have been reported in Section 4.8.

4.2 A Brief Literature Review

The literature on growth convergence initiated by the seminal papers of Solow (1956) and Swan (1956) is vast and it reached the 'formal specification' stage with the influential work of Barro and Salai-Martin (1992) (henceforth BS) and Mankiw, Romer and Weil (1992), (henceforth MRW), which derived the regression specification from the neoclassical growth model. MRW is based on original Solow-Swan model, and BS on Cass-Koopmans' (1965) optimal savings model. Both papers derive the law of motion of capital and income around the steady state and then translate that into an estimable *cross-sectional regression* equation. Similar results on conditional convergence across countries are presented in Holtz-Eakin (1993), Sala-i-Martin (1996) and many others. One of the crucial assumption of these cross-sectional models is that the differences in initial unobserved technology diffusion is considered to be a part of error terms. This assumption makes their equations estimable by ordinary least square (OLS) method. Thus the cross-section approach to convergence encountered some important limitations. Temple (1998) discussed the influence of possible measurement errors on the results of MRW. The basic limitation lies in the fact that having just one data for a country provides a weak basis for estimation of the convergence, which refers primarily to a within-country process. There is too much heterogeneity across countries to validate the assumption that cross country data can be treated as multiple data of the same country. Thus, the convergence research gradually moved to other approaches like time-series and panel methods.

Lee, Pesaran and Smith (1992), Quah (1992), Binder and Pesaran (1999) support for *time series regression* for each country separately to analyze the conditional convergence hypothesis. In simple terms, convergence using time-series approach, implies whether income of a specific country has unit root or not. They argue that standard cross-section methods throw away useful information which can be taken care by analyzing each country separately. Moreover, time series analysis has been applied to investigate across convergence too, see for instance Quah (1992), Bernard and Durlauf (1995) and Evans and Karras (1996). Broadly speaking the time series analysis supports a variant of conditional convergence hypothesis and thus results are not much different from those implied by cross-section methods.

One of the crucial limitation of the cross-section approach is that it cannot capture the technological diffusion and capital deepening process, which are vital for income convergence across countries. Thus many researchers used *panel methods* to capture such technology diffusion by introducing individual effects in the model. However, there are many ways to model the *country-specific effect*. For instance, Islam (1995) strongly supports fixed effect estimation due to the assumption of correlation of unobserved technology diffusion with the regressors. The key strength of this method is that it takes care of one form of heterogeneity: any omitted variables that are constant over time will not bias the estimates, even when the omitted variables are correlated with the explanatory variables.

There are, however, some concerns about fixed effect specification. For instance, sometimes a variable of interest is measured at only one point in time, and even if the variables are measured at more frequent intervals, they are sometimes highly persistent. In that case the within-country variation is unlikely to be informative. Too often researchers use fixed effects to analyze the effects of variables that are fairly constant over time, or that affect growth only with a long time lag. Standard transformation like first differences or "within groups" transformations are likely to exacerbate the problem of measurement errors. They lead to large reduction in precision of the parameter estimates since the between-country variation is thrown away. Koop and Steel (2000) argues that much of variation in efficiency level occurs *between* rather than *within* countries. Thus, a random effects generalized least square (GLS) estimator will be more efficient than within-country estimator when the random effects assumptions are appropriate. Durlauf and Quah (1999) point out that the individual effects are of fundamental interest to growth economists as they appear to be the key source of persistent income differences. Thus they advocate for modeling the heterogeneity in the model rather than finding the ways to eliminate its effects. In this paper, I adopt a random effects model as I intend to *test* the significance of individual effects in the presence of time dynamics and spatial dependencies, rather than just treating them as the nuisance parameters, as is done in fixed effects model.

Recently, many researchers are using *spatial models* to analyze growth convergence. It is a known fact that the economies are assumed to be independent in the neoclassical growth theory. However, with globalization, technological advances in one economy are transmitted to other economies. Thus, the closed independent economy assumption are not valid, and one needs to take into account the possible spatial correlation, both in cross-sectional and panel data settings. From statistical point of view, ignoring the presence of spatial dependence leads to unreliable inference. In recent years many researchers have used spatial methods to capture such technology interdependence and knowledge spillover effects. The main idea is to capture the impact of cross-country spillovers on growth process. There are many ways to measure this interdependence. One of the most common way is to express the aggregate level of technology of any country to be dependent on the stock of knowledge/capital of its neighbors or trading partners by using geographic and economic distances. For instance, Conley and Ligon (2002), Ertur, Gallo and Baumont (2006), Ertur and Koch (2007), Yu and Lee (2009) and many others, have used spatial approach to analyze growth convergence.

Each method, as I discussed, has its own merits and drawbacks. It is evident from the discussion that the convergence research has not produced any concrete consensus. Given the differences in approach, sample, data, model, estimation technique, etc., absence of consensus is not surprising though. The crucial issue is to find a good approximation to DGP, and to achieve this objective, I start out with a general model in the following section.

4.3 The Model Setup

The model setup is the combination of all the piecewise frameworks I discussed earlier:

$$y_{it} = \gamma y_{it-1} + \tau \sum_{j=1}^{N} m_{ij} y_{jt} + \delta \sum_{j=1}^{N} m_{ij} y_{jt-1} + X_{it}\beta + u_{it} \tag{4.1}$$

$$u_{it} = \mu_i + \epsilon_{it} \tag{4.2}$$

$$\epsilon_{it} = \lambda \sum_{j=1}^{N} w_{ij} \epsilon_{jt} + v_{it} \tag{4.3}$$

$$v_{it} = \rho v_{it-1} + e_{it}, \text{where } e_{it} \sim IIDN(0, \sigma_e^2) \tag{4.4}$$

for $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$. Here y_{it} is the observation for i^{th} location/unit at t^{th} time, X_{it} denotes the observations on non-stochastic regressors and e_{it} is the regression disturbance. Spatial dependence is captured by the weight matrices $M = (m_{ij})$ and $W = (w_{ij})$. Here m_{ij} and w_{ij} are the (i, j) th element of weight matrices M and W respectively, which capture the interdependence of income and unobserved error terms between the country i and j.The matrices M and W are each row-standardized and the diagonal elements are set to zero. In this model framework, time dynamics (γ), random effects (μ_i) with $\mu_i \sim IID(0, \sigma_\mu)$, serial correlation ($\rho$), space recursive ($\delta$), spatial lag dependence (τ) and spatial error dependence

(λ) are considered.

In matrix form, the equations (4.1) - (4.4) can be written compactly as

$$y = \tau(I_T \otimes M)y + [(\gamma + \delta M) \otimes I_T]ly + X\beta + u, \tag{4.5}$$

where y is of dimension $NT \times 1, X$ is $NT \times K, \beta$ is $k \times 1$, u is $NT \times 1$, I_T is an identity matrix of dimension $T \times T$ and $\otimes$ denotes Kronecker product. Here l is the lag operator, X is assumed to be of full column rank and its elements are bounded in absolute value. The disturbance term can be expressed as

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})v. \tag{4.6}$$

Here $B = (I_N - \lambda W)$ and ι_T is vector of ones of dimension T. Under this setup, the variance-covariance matrix of u is given by

$$\Omega = \sigma_{\mu^2}[J_T \otimes I_N] + [V \otimes (B'B)^{-1}], \tag{4.7}$$

where J_T is a matrix of ones of dimension $T \times T$, and V is the familiar $T \times T$ variance -covariance matrix for AR (1) process in equation (4.4), i.e.,

$$V = E(v'v) = [\frac{1}{1-\rho^2}V_1] \otimes \sigma_e^2 I_N = V_\rho \otimes \sigma_e^2 I_N, \tag{4.8}$$

with

$$V_1 = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & 1 \end{bmatrix},$$

and $V_\rho = \frac{1}{1-\rho^2}V_1$.

The log-likelihood function of the above model can be written as:

$$L = \frac{-NT}{2}ln2\pi - \frac{1}{2}\ln|\Omega| + T\ln|A| -$$
$$\frac{1}{2}[(I_T \otimes A)y - [(\gamma + \delta M) \otimes I_T]ly - X\beta]'\Omega^{-1}[(I_T \otimes A)y - [(\gamma + \delta M) \otimes I_T]ly - X\beta] \tag{4.9}$$

where $A = (I_N - \tau M)$. Following Chapter 3, I can write

$$\frac{1}{2}\ln|\Omega| = -\frac{N}{2}\ln(1-\rho^2) + \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| + \frac{NT}{2}\ln\sigma_e^2 - (T-1)\ln|B|,$$

where $d^2 = \alpha^2 + (T-1)$, $\alpha = \sqrt{\frac{1+\rho}{1-\rho}}$ and $\phi = \frac{\sigma_\mu^2}{\sigma_e^2}$. Substituting $\frac{1}{2}\ln|\Omega|$ in L , I obtain

$$\begin{aligned} L = &\frac{-NT}{2} ln2\pi + \frac{N}{2}\ln(1-\rho^2) - \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| - \frac{NT}{2}\ln\sigma_e^2 + (T-1)\ln|B| + T\ln|A| \\ &- \frac{1}{2}[(I_T\otimes A)y - [(\gamma+\delta M)\otimes I_T]ly - X\beta]'\Omega^{-1}[(I_T\otimes A)y - [(\gamma+\delta M)\otimes I_T]ly - X\beta] \end{aligned} \tag{4.10}$$

The likelihood function in equation (4.10), will be used to derive the test statistics in the next section. Now I state the assumptions, needed for the validity of the asymptotic properties.

Assumption 1.
(i) W and M are row-standardized weight matrices whose diagonal elements are zero.
(ii) W and M are uniformly bounded in row and column sums in absolute value and $(I - \lambda W)^{-1}$ and $(I - \tau M)^{-1}$ are also uniformly bounded.
(iii) $A_n = (I - \tau M)^{-1}(\gamma I + \delta M)$ is also uniformly bounded in row and column sums in absolute value.

Assumption 2. The disturbances $e_{it}, i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, are *i.i.d.* across i and t with zero mean, variance σ_e^2 and $E|e_{it}|^{4+\eta} < \infty$, for some $\eta > 0$.

Assumption 3. The element of X_{Nt} are nonstochastic and bounded uniformly (BU) in n and T. Also, $\lim_{T\to\infty} \frac{1}{NT}\sum_{t=1}^{T} X'_{Nt}X_{Nt}$ exists and is nonsingular.

Assumption 4. N is a nondecreasing function of T and T goes to infinity.

Assumption 1 is standard assumption in spatial analysis and boundedness condition on W,

M and $(I - \lambda W)$, $(I - \tau M)$ and A_n, limit the spatial correlation of manageable degree. *Assumption 2* provides regularity assumptions on for e_{it}. When exogenous variables X_{Nt} are included in the model, it is convenient to assume that they are uniformly bounded as in *Assumption 3*. Lastly, if *Assumption 4* holds then we can say that $N, T \to \infty$ simultaneously.

4.4 Derivation of the Test Statistics

The full model (4.1) - (4.4) has the following: linear regression coefficients and innovation variance (β, σ_e^2), time dynamics γ, random effects σ_μ^2, time-series correlation ρ, space-time dynamics parameter δ, spatial error dependence λ, and spatial lag dependence τ. The full parameter vector will be denoted by $\theta = (\beta', \sigma_e^2, \gamma, \sigma_\mu^2, \rho, \delta, \lambda, \tau)'$. I am interested in testing significance of last six parameters individually in the possible presence of the rest. For example, in order to detect the time- dynamics I would test, say, $H_o^b : \gamma = 0$ in presence of $\phi = (\sigma_\mu^2, \rho, \delta, \tau, \lambda)'$. The usual practice is using likelihood ratio test and conditional RS tests. However, those tests require estimation of both γ and ϕ (or ϕ alone) along with (β, σ_e^2). Instead, in this paper I contribute to the existing literature by developing adjusted RS tests for specification search in dynamic panel spatial framework by using Bera and Yoon (1993) test principle, which requires estimation of the simplest model under joint null of no misspecification.

All the proposed adjusted tests are based on the joint null hypothesis (of no misspecification), i.e., $H_0^a : \gamma = \sigma_\mu^2 = \rho = \delta = \lambda = \tau = 0$. Thus under H_o^a, the parameter vector is $\theta_o = (\beta', \sigma_e^2, 0, 0, 0, 0, 0, 0)'$. The proposed tests can take care of the possible presence of all the nuisance parameters indirectly through their respective Fisher-Rao score evaluated under the joint null. Due to this estimation simplicity, the suggested tests are more amenable to use by empirical researchers than the LR or conditional RS tests.

4.4.1 Bera and Yoon Test Principle

Consider a general model represented by the log-likelihood $L(\omega, \psi, \phi)$ where the parameters ω, ψ and ϕ are, respectively, $(p\times 1), (r\times 1)$ and $(s\times 1)$ vectors. Here I assume that underlying density function satisfies the regularity conditions, as stated in Serfling (1980), Lehmann and Romano (2005), for the MLE to have asymptotic Gaussian distribution. Suppose a researcher sets $\phi = \phi_0$ and tests $H_0 : \psi = \psi_0$ using the log-likelihood function $L_1(\omega, \psi) = L(\omega, \psi, \phi_0)$, where ψ_0 and ϕ_0 are known. The RS statistic for testing H_0 in $L_1(\omega, \psi)$ will be denoted by RS_ψ. Let us denote $\theta = (\omega', \psi', \phi')'_{(p+r+s)\times 1}$ and $\tilde{\theta} = (\tilde{\omega}', \psi_0', \phi_0')'_{(p+r+s)\times 1}$, where $\tilde{\omega}$ is the ML estimator of ω under $\psi = \psi_0$ and $\phi = \phi_0$. I define the score vector and the information matrix, respectively, as

$$d_a(\theta) = \frac{\partial L(\theta)}{\partial a} \qquad \text{and} \qquad J(\theta) = -E[\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}] = \begin{bmatrix} J_\omega & J_{\omega\psi} & J_{\omega\phi} \\ J_{\psi\omega} & J_\psi & J_{\psi\phi} \\ J_{\phi\omega} & J_{\phi\psi} & J_\phi \end{bmatrix} \tag{4.11}$$

where $a = (\omega, \psi, \phi)$ and n is the sample size. If $L_1(\omega, \psi)$ were the true model, then it is well known that under $H_0 : \psi = \psi_0$

$$RS_\psi = \frac{1}{n} d_\psi(\tilde{\theta}) J_{\psi.\omega}(\tilde{\theta})^{-1} d_\psi(\tilde{\theta})' \to \chi^2_r(0) \tag{4.12}$$

where $J_{\psi.\omega}(\tilde{\theta}) = J_\psi - J_{\psi\omega} J_\omega^{-1} J_{\omega\psi}$.

Under local alternative $H_1 : \psi = \psi_0 + \frac{\zeta}{\sqrt{n}}$, $RS_\psi \to \chi^2_r(\lambda_1)$, where the non-centrality parameter $\lambda_1 \equiv \lambda_1(\zeta) = \zeta' J_{\psi.\omega}\zeta$. Given this setup, i.e., under no misspecification, asymptotically the test will have the correct size and locally optimal. Now suppose that the true log-likelihood function is $L_2(\omega, \phi)$ so that the considered alternative $L_1(\omega, \psi)$ is (completely) misspecified. Using the local misspecification $\phi = \phi_0 + \frac{\delta}{\sqrt{n}}$, Davidson and MacKinnon (1987) and Saikkonen (1989) derived the asymptotic distribution of RS_ψ under $L_2(\omega, \phi)$ as $RS_\psi \to \chi^2_r(\lambda_2)$, where the non-centrality parameter $\lambda_2(\delta) = \delta' J_{\phi\psi.\omega} J^{-1}_{\psi.\omega} J_{\psi\phi.\omega}\delta$ with $J_{\psi\phi.\omega} = J_{\phi\psi} - J_{\phi\omega} J_\omega^{-1} J_{\omega\psi}$. Owing to the presence of this non-centrality parameter λ_2, RS_ψ will reject the true null hypothesis $H_0 : \psi = \psi_0$ more often, i.e., the test will have excessive size. Here the crucial term is $J_{\phi\psi.\omega}$ which can be interpreted as partial covariance between the score vectors d_ϕ and d_ψ after eliminating the linear effect of d_ω on d_ϕ and d_ψ. If $J_{\psi\phi.\omega} = 0$, then asymptotically the local presence of ϕ has no effect on RS_ψ. Bera and Yoon (1993) suggested a modification to RS_ψ to overcome this problem of over-rejection, so that the resulting test is valid under the local presence of ϕ. The modified statistic is given by

$$RS_{\psi}^{*} = \frac{1}{N}[d_{\psi}(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})d_{\phi}(\tilde{\theta})]'[J_{\psi.\omega}(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})J_{\phi\psi.\omega}(\tilde{\theta})]^{-1}$$
$$[d_{\psi}(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})d_{\phi}(\tilde{\theta})]'. \quad (4.13)$$

This new test essentially adjusts the mean and variance of the standard RS statistics RS_{ψ}, and, under $H_o : \psi = \psi_o$

$$RS_{\psi}^{*} \to \chi_r^2(0) \quad (4.14)$$

while under $H_1 : \psi = \psi_0 + \frac{\zeta}{n}$,

$$RS_{\psi}^{*} \to \chi_r^2(\lambda_3) \quad (4.15)$$

where $\lambda_3 = \zeta'(J_{\psi.\omega} - J_{\psi\phi.\omega}J_{\psi.\omega}^{-1}J_{\phi\psi.\omega})\zeta$. Note the results in (14) and (15) are valid both under presence or absence of local misspecification, since the asymptotic distribution of RS_{ψ}^{*} is unaffected by the local departure of ϕ from ϕ_0.

BY shows that for local misspecification the adjusted test is asymptotically equivalent to Neyman's $C(\alpha)$ test and thus shares its optimal properties. Three observations are worth noting regarding RS_{ψ}^{*}. First, RS_{ψ}^{*} requires estimation only under the joint null, namely $\psi = \psi_0$ and $\phi = \phi_0$. That means, in most cases, as we will see later, we can conduct our tests based on only OLS residuals. Given the full specification of the model $L(\omega, \psi, \phi)$, it is of course possible to derive RS test for $\psi = \psi_0$ after estimating ϕ (and ω) by MLE, which are generally referred to as conditional tests. However, ML estimation of ϕ could be difficult in some instances. Second, when $J_{\psi\phi.\omega} = 0, RS_{\psi}^{*} = RS_{\psi}$, which is a simple condition to check. If this condition is true, RS_{ψ} is an asymptotically valid test in the local presence of ϕ. Finally, let $RS_{\psi\phi}$ denote the joint RS test statistic for testing hypothesis of the form $H_0 : \psi = \psi_0$ and $\phi = \phi_0$ using the alternative model $L(\omega, \psi, \phi)$. Then it be shown that [for a proof see Bera, Bilias and Yoon (2007), Bera, Montes-Rojas and Sosa-Escudero (2009)]

$$RS_{\psi\phi} = RS_{\psi}^{*} + RS_{\phi} = RS_{\phi}^{*} + RS_{\psi}, \quad (4.16)$$

where RS_{ϕ} and RS_{ϕ}^{*} are, respectively, the counterparts of RS_{ψ} and RS_{ψ}^{*} for testing H_0 :

$\phi = \phi_0$. This is a very important identity since it implies that a joint RS test for two parameter vectors ψ and ϕ can be decomposed into sum of two orthogonal components: (i) the adjusted statistic for one parameter vector and (ii) (unadjusted) marginal test statistic for the other. Since many econometrics softwares provide the marginal (and sometime the joint) test statistics, the adjusted versions can be obtained effortlessly.

Significance of $RS_{\psi\phi}$ indicates some form of misspecification in the basic model with parameter ω only. However, the correct source(s) of departure can be identified only by using the adjusted statistics RS^*_ψ and RS^*_ϕ not the marginal ones (RS_ψ and RS_ϕ). This testing strategy is close to the idea of Hillier (1991) in the sense that it partitions the overall rejection region to obtain evidence about the specific direction(s) in which the basic model needs revision. And it achieves that without estimating any of the nuisance parameters. For detailed discussion, see Sen and Bera (2011).

4.4.2 Score Functions and Information Matrix

Recall that for the dynamic panel spatial model, the full parameter vector was $\theta = (\beta', \sigma_e^2, \gamma, \sigma_\mu^2, \rho, \delta, \lambda, \tau)'$. In context the notation of Section 4.4.1, $\theta = (\omega', \psi', \phi'),'$ with $\omega = (\beta', \sigma_e^2)$ and ψ and ϕ could be any combinations of the parameters under test, namely $(\gamma, \sigma_\mu^2, \rho, \delta, \lambda, \tau)$. The restricted model (under the joint null, $H_0^a : \gamma = \sigma_\mu^2 = \rho = \delta = \lambda = \tau = 0$) is simple panel model, i.e., $y_{it} = X_{it}\beta + u_{it}$, where, $u_{it} \sim IIDN(0, \sigma_u^2)$. For simplicity I assume the weight matrices W and M to be same, which is often realistic in practice. The score functions and information matrix J evaluated at the restricted MLE of θ with $\tilde{\omega} = (\tilde{\beta}, \tilde{\sigma_e^2})$ are:

$$\frac{\partial L}{\partial \beta} = 0 \tag{4.17}$$

$$\frac{\partial L}{\partial \sigma_e^2} = 0 \tag{4.18}$$

$$\frac{\partial L}{\partial \gamma} = \frac{\tilde{u}'[I_T \otimes Y_{NT-1}]}{\sigma_e^2} \tag{4.19}$$

$$\frac{\partial L}{\partial \sigma_\mu^2} = \frac{NT}{2\tilde{\sigma_e^2}}[\frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1] \tag{4.20}$$

$$\frac{\partial L}{\partial \rho} = \frac{NT}{2}[\frac{\tilde{u}'(G \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{4.21}$$

$$\frac{\partial L}{\partial \delta} = \frac{\tilde{u}'[(I_T \otimes W)Y_{NT-1}]}{\tilde{\sigma_e^2}} \tag{4.22}$$

$$\frac{\partial L}{\partial \tau} = \frac{\tilde{u}'[(I_T \otimes W)Y_{NT}]}{\tilde{\sigma_e^2}} \tag{4.23}$$

$$\frac{\partial L}{\partial \lambda} = \frac{NT}{2}[\frac{\tilde{u}'(I_T \otimes (W + W'))\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{4.24}$$

where $\tilde{u} = y - x\tilde{\beta}$ is the OLS residual vector of dimension $NT \times 1$, $\tilde{\sigma_e^2} = \frac{\tilde{u}'\tilde{u}}{NT}$ and $G = \frac{\partial V_1}{\partial \rho}|_{H_o^a}$, where G is bidiagonal matrix with bidiagonal elements all equal to one. Y_{NT} and Y_{NT-1} are vector of y and lagged values of y respectively, each of dimension $NT \times 1$.

The information matrix J, equation (4.11), under H_o^a is

$$J(\theta_o) = \begin{bmatrix} J_\beta & 0 & J_{\beta\gamma} & 0 & 0 & J_{\beta\delta} & J_{\beta\tau} & 0 \\ 0 & J_{\sigma_e^2} & J_{\sigma_e^2\gamma} & J_{\sigma_e^2\sigma_\mu^2} & 0 & 0 & 0 & 0 \\ J_{\gamma\beta} & J_{\gamma\sigma_e^2} & J_\gamma & J_{\gamma\sigma_\mu^2} & J_{\gamma\rho} & 0 & 0 & 0 \\ 0 & J_{\sigma_\mu^2\sigma_e^2} & J_{\sigma_\mu^2\gamma} & J_{\sigma_\mu^2} & J_{\sigma_\mu^2\rho} & 0 & 0 & 0 \\ 0 & 0 & J_{\rho\gamma} & J_{\rho\sigma_\mu^2} & J_\rho & 0 & 0 & 0 \\ J_{\delta\beta} & 0 & 0 & 0 & 0 & J_\delta & J_{\delta\tau} & J_{\delta\lambda} \\ J_{\tau\beta} & 0 & 0 & 0 & 0 & J_{\tau\delta} & J_\tau & J_{\tau\lambda} \\ 0 & 0 & 0 & 0 & 0 & J_{\lambda\delta} & J_{\lambda\tau} & J_\lambda \end{bmatrix} \tag{4.25}$$

where $J = E(-\frac{1}{NT}\frac{\partial^2 L}{\partial\tilde{\theta}\partial\tilde{\theta}'})$ evaluated at θ_0. The detailed derivation and expression of each of the terms of the information matrix are relegated to the appendix B.

Apart from the RS statistic for full joint null hypothesis H_0^a, I propose *six* (modified) test statistics for the following hypotheses:

I) $H_o^b : \gamma = 0$ in presence of $\sigma_\mu^2, \rho, \delta, \tau, \lambda$.

II) $H_o^c : \sigma_\mu^2 = 0$ in presence of $\gamma, \rho, \delta, \tau, \lambda$.

III) $H_o^d : \rho = 0$ in presence of $\gamma, \sigma_\mu^2, \delta, \tau, \lambda$.

IV) $H_o^e : \delta = 0$ in presence of $\gamma, \sigma_\mu^2, \rho, \tau, \lambda$.

V) $H_o^f : \tau = 0$ in presence of $\gamma, \sigma_\mu^2, \rho, \delta, \lambda$.

VI) $H_o^g : \lambda = 0$ in presence of $\gamma, \sigma_\mu^2, \rho, \delta, \tau$.

Decision on testing these six hypotheses can guide a researcher to identify the correct source(s) of departure(s) from H_o^a when it is rejected. One can test various combinations of I) to VI) by testing on two/three/four parameters at a time under the null and compute additional *ninety* test statistics ($C_2^6 + C_3^6 + C_4^6 = 90$). I will demonstrate that it is unnecessary as these six (I -VI) tests are "sufficient" to detect any misspecification in the basic model. Also keeping the total number of tests to a minimum is beneficial to avoid the pre-testing problem. Since by construction the proposed tests are independent of each other, so one can easily compute the overall significance level.

Given the full model specification, it is of-course possible to derive conditional RS and LR tests, say for $\lambda = 0$ in presence of $\gamma, \sigma_\mu^2, \rho, \delta$ and τ, however that would entail the estimation of $\gamma, \sigma_\mu^2, \rho, \delta$ and τ parameters and also of λ for LR test. For the adjusted RS test these estimations are not required as it indirectly takes care of the possible presence of nuisance parameters through the Fisher-Rao score function.

Let us take the case for $H_o^g : \lambda = 0$ in presence of $\phi = (\gamma, \sigma_\mu^2, \rho, \delta, \tau)$, i.e., testing for spatial error dependence in presence of time dynamics (γ), random effect (σ_μ^2), serial correlation (ρ), space-time dynamics (δ) and spatial lag dependence (τ). For this hypothesis, the term $J_{\psi\phi.\omega}$, i.e., $J_{\lambda\phi.\omega} \neq 0$ where $\phi = (\gamma, \sigma_\mu^2, \rho, \delta, \tau)'$ and $\omega = (\beta', \sigma_e^2)'$. The term $J_{\lambda\phi.\omega}$ can be interpreted as partial covariance of scores of λ and ϕ after eliminating the linear effect of ω. Therefore, the parameter λ is not "independent" of $(\gamma, \sigma_\mu^2, \rho, \delta, \tau)'$ and vice versa. Thus, the marginal RS test statistic based on the score d_λ, i.e., RS_λ for $H_0^b : \lambda = 0$ assuming $\phi = (\gamma, \sigma_\mu^2, \rho, \delta, \tau) = (0, 0, 0, 0, 0)$ is not valid test under the presence of ϕ. In the next subsection, the proposed test statistic, RS_λ^*, that eliminates the effects of ϕ without estimating them, and is more appropriate test will be presented. I will further show, that test statistic for λ is dependent on δ and τ only, i.e., it is asymptotically independent of $(\gamma, \sigma_\mu^2, \rho)$. Thus even if one is interested to evaluate conditional RS test for H_o^g , then estimation of all the parameters are not necessary as the test statistic for λ in presence of ϕ is only affected by the presence of other spatial parameters, i.e., δ and τ and not by *panel* parameters. These type of analysis of inter-dependencies cannot be done using LR tests. The proposed adjusted tests make this possible in an elegant way. Of course, given the current computing power, it is not difficult to estimate a complex model; however, it could be sometime hard to ensure the stability of many parameter estimates. Also theoretically the stationarity regions of the parameter space have not been fully worked for the spatial model [See Elhorst 2010].

4.4.3 Adjusted RS tests

In this section I present the adjusted test statistics one-by-one, each of which asymptotically follows χ^2_1 distribution under the respective null hypothesis. Detailed derivation are in the appendix.

I) $H^b_o : \gamma = 0$ in presence of $\phi = (\sigma^2_\mu, \rho, \delta, \tau, \lambda)$.

To recall, here I am testing the significance of time-dynamics (γ) in presence of random effects (σ^2_μ), serial correlation (ρ), and spatial dependence (δ, τ, λ). Using the information matrix in (25), $J_{\gamma\phi.\omega}$, which can be interpreted as a covariance between parameter of interest, i.e., γ and rest of the parameters, i.e., $\phi = (\sigma^2_\mu, \rho, \delta, \tau, \lambda)$ is given by $J_{\gamma\phi.\omega} = (J_{\gamma\sigma^2_\mu.\sigma^2_e}, J_{\gamma\rho}, 0, 0, 0)$. From this we can infer:

(i) unadjusted RS test for H^b_o is not valid.

(ii) the partial covariance of d_γ and $d_{\sigma^2_\mu}$, and, d_γ and d_ρ are nonzero; while covariance with the spatial parameters (δ, τ, λ) are zero. Thus the test for γ is affected by the presence of σ^2_μ and ρ only.

The adjusted test, RS^*_γ, takes care of the presence of σ^2_μ and ρ using the score function of σ^2_μ and ρ, i.e., equations (3.20) and (3.21). These scores can be viewed as "sufficient" statistics and thus can be interpreted as the indirect estimators of the respective parameters. For example, in a simple time-series model, $\hat{\rho} = \frac{\sum \tilde{u}_t \tilde{u_{t-1}}}{\sum \tilde{u}_t \tilde{u}_t{}'}$, and Durbin-Watson test, which is essentially a RS test, are related by: DW $\approx 2(1 - \hat{\rho})$. Here, instead of direct estimation of the nuisance parameters, ρ and σ^2_μ , the adjusted test utilizes their respective scores, i.e.,

$$d_{\sigma^2_\mu} = \frac{\partial L}{\partial \sigma^2_\mu} = \frac{NT}{2\tilde{\sigma^2_e}}[\frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1]$$

$$d_\rho = \frac{\partial L}{\partial \rho} = \frac{NT}{2}[\frac{\tilde{u}'(G \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}}]$$

One can note the similarity between $\hat{\rho}$ and d_ρ. The adjusted test statistic for $H_o : \gamma = 0$ is:

$$RS^*_\gamma = \frac{[d_\gamma - J_{\gamma\sigma^2_\mu.\sigma^2_e} J^{-1}_{\sigma^2_\mu.\sigma^2_e} d_{\sigma^2_\mu} - J_{\gamma\rho} J^{-1}_\rho d_\rho]^2}{[J_{\gamma.\omega} - J_{\gamma\sigma^2_\mu.\sigma^2_e} J^{-1}_{\sigma^2_\mu.\sigma^2_e} J_{\sigma^2_\mu\gamma.\sigma^2_e} - J_{\gamma\rho} J^{-1}_\rho J_{\rho\gamma}]}. \tag{4.26}$$

While the unadjusted counterpart is

$$RS_\gamma = \frac{d^2_\gamma}{J_{\gamma.\omega}}. \tag{4.27}$$

From equation (4.26) it is quite apparent how RS^*_γ takes care of the possible presence of nuisance parameters (σ^2_μ, ρ) using their respective scores evaluated under joint null H^a_o. It is based on the effective score $d^*_\gamma = [d_\gamma - J_{\gamma\sigma^2_\mu.\sigma^2_e} J^{-1}_{\sigma^2_\mu.\sigma^2_e} d_{\sigma^2_\mu} J_{\gamma\rho} J^{-1}_\rho d_\rho]$ which renders d^*_γ to be orthogonal to $d_{\sigma^2_\mu}$ and d_ρ. For other nuisance parameters, (δ, τ, λ), no such adjustments are necessary as it is evident from the expression of $J_{\gamma\phi.\omega}$ that $J_{\gamma(\delta\tau\lambda)} = (0, 0, 0)$, i.e., asymptotically they do not affect γ as far as testing is concerned. Thus inference on γ is affected only by the presence of panel and time-series parameters i.e., σ^2_μ and ρ, and *not* by the presence of any spatial parameters (δ, τ, λ). This separation between time and space parameters is quite interesting, and RS^*_γ takes full advantage of it which is not possible under other test procedures.

For the following hypotheses, I mention the respective test statistics.

II) $H^c_o : \sigma^2_\mu = 0$ in presence of $\gamma, \rho, \delta, \tau, \lambda$.

Here, I am testing for random effects (σ^2_μ), in presence of time dynamics (γ), serial correlation of errors (ρ), space-time dynamics (δ), spatial lag dependence (τ) and spatial error dependence (λ). The crucial quantity is $J_{\sigma^2_\mu\phi.\omega} = (J_{\sigma^2_\mu\gamma.\sigma^2_e}, J_{\sigma^2_\mu\rho}, 0, 0, 0)$ where $\phi = (\gamma, \rho, \delta, \tau, \lambda)$

The adjusted RS test statistics is:

$$RS^*_{\sigma^2_\mu} = \frac{[d_{\sigma^2_\mu} - J_{\sigma^2_\mu\gamma.\sigma^2_e} J^{-1}_{\gamma.\omega} d_\gamma - J_{\sigma^2_\mu\rho} J^{-1}_\rho d_\rho]^2}{[J_{\sigma^2_\mu.\sigma^2_e} - J_{\sigma^2_\mu\gamma.\sigma^2_e} J^{-1}_{\gamma.\omega} J_{\gamma\sigma^2_\mu.\sigma^2_e} - J_{\sigma^2_\mu\rho} J^{-1}_\rho J_{\rho\sigma^2_\mu}]}, \tag{4.28}$$

and the unadjusted one

$$RS_{\sigma_\mu^2} = \frac{d_{\sigma_\mu^2}^2}{J_{\sigma_\mu^2.\sigma_e^2}} \tag{4.29}$$

Here σ_μ^2 is dependent only on γ and ρ, therefore $RS_{\sigma_\mu^2}$ uses the effective score $d^*_{\sigma_\mu^2} = [d_{\sigma_\mu^2} - J_{\sigma_\mu^2\gamma.\sigma_e^2} J_{\gamma.\omega}^{-1} d_\gamma - J_{\sigma_\mu^2\rho} J_\rho^{-1} d_\rho]$, making it orthogonal to d_γ and d_ρ.

III) $H_o^d : \rho = 0$ in presence of $\gamma, \sigma_\mu^2, \delta, \tau, \lambda$.

Here, $\phi = (\gamma, \sigma_\mu^2, \delta, \tau, \lambda)$ and $J_{\rho\phi.\omega} = (J_{\rho\gamma}, J_{\rho\sigma_\mu^2}, 0, 0, 0)$.

Thus, the adjusted test statistic is:

$$RS_\rho^* = \frac{[d_\rho - J_{\rho\gamma.\sigma_e^2} J_{\gamma.\omega}^{-1} d_\gamma - J_{\rho\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2.\sigma_e^2}^{-1} d_{\sigma_\mu^2}]^2}{[J_\rho - J_{\rho\gamma.\sigma_e^2} J_{\gamma.\omega}^{-1} J_{\gamma\rho.\sigma_e^2} - J_{\rho\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2.\sigma_e^2}^{-1} J_{\sigma_\mu^2\rho.\sigma_e^2}]} \tag{4.30}$$

and while the unadjusted one is :

$$RS_\rho = \frac{d_\rho^2}{J_\rho} \tag{4.31}$$

It is evident from the term $J_{\rho\phi.\omega}$, that among all the nuisance parameters, serial correlation (ρ) is directly affected by presence of only time dynamics (γ) and random effects (σ_μ^2). Thus, the effective score of the test statistic $[d_\rho^* = d_\rho - J_{\rho\gamma.\sigma_e^2} J_{\gamma.\omega}^{-1} d_\gamma - J_{\rho\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2.\sigma_e^2}^{-1} d_{\sigma_\mu^2}]$, in equation (30), clearly reveals this, i.e., d_ρ^* is made orthogonal to d_γ and $d_{\sigma_\mu^2}$.

IV) $H_o^e : \delta = 0$ in presence of $\gamma, \sigma_\mu^2, \rho, \tau, \lambda$.

Here, $\phi = (\gamma, \sigma_\mu^2, \rho, \tau, \lambda)$ and $J_{\delta\phi.\omega} = (0, 0, 0, J_{\delta\lambda.\beta}, J_{\delta\tau.\beta})$.

The adjusted RS test statistic is:

$$RS_\delta^* = \frac{[d_\delta - J_{\delta\lambda.\beta} J_{\lambda.\beta}^{-1} d_\lambda - J_{\delta\tau.\beta} J_{\tau.\beta}^{-1} d_\tau]^2}{[J_{\delta.\beta} - J_{\delta\lambda.\beta} J_{\lambda\beta}^{-1} J_{\lambda\delta.\beta} - J_{\delta\tau.\beta} J_{\tau.\beta}^{-1} J_{\tau\delta.\beta}]} \tag{4.32}$$

and the unadjusted one is

$$RS_{\delta} = \frac{d_{\delta}^2}{J_{\delta.\beta}} \tag{4.33}$$

V) $H_o^f : \tau = 0$ in presence of $\gamma, \sigma_{\mu}^2, \rho, \delta, \lambda$.

The set of nuisance parameters is $\phi = (\gamma, \sigma_{\mu}^2, \rho, \delta, \lambda)$, and $J_{\tau\phi.\omega} = (0, 0, 0, J_{\tau\delta.\beta}, J_{\tau\lambda.\beta})$.

The adjusted test statistic is:

$$RS_{\tau}^{*} = \frac{[d_{\tau} - J_{\tau\delta.\beta} J_{\delta.\beta}^{-1} d_{\delta} - J_{\tau\lambda.\beta} J_{\lambda.\beta}^{-1} d_{\lambda}]^2}{[J_{\tau.\beta} - J_{\tau\delta.\beta} J_{\delta.\beta}^{-1} J_{\delta\tau.\beta} - J_{\tau\lambda.\beta} J_{\lambda.\beta}^{-1} J_{\lambda\tau.\beta}]} \tag{4.34}$$

The unadjusted test statistic is:

$$RS_{\tau} = \frac{d_{\tau}^2}{J_{\tau.\beta}} \tag{4.35}$$

Similar to the other proposed test statistics, RS_{τ}^{*} also takes care of the presence of nuisance parameters through their respective scores. This is evident from the equation (3.34).

Lastly,

VI) $H_o^g : \lambda = 0$ in presence of $\gamma, \sigma_{\mu}^2, \rho, \delta, \tau$.

Here, $\phi = (\gamma, \sigma_{\mu}^2, \rho, \delta, \tau)$ and the partial covariance term, i.e., $J_{\lambda\phi.\omega} = (0, 0, 0, J_{\lambda\delta}, J_{\lambda\tau})$. The proposed adjusted test takes care of the interdependence of the two nuisance parameters, space-time dynamics (δ) and spatial lag dependence (τ), through their score functions. Thus, the adjusted test statistic is:

$$RS_{\lambda}^{*} = \frac{[d_{\lambda} - J_{\lambda\delta} J_{\delta.\beta}^{-1} d_{\delta} - J_{\lambda\tau} J_{\tau.\beta}^{-1} d_{\tau}]^2}{[J_{\lambda} - J_{\lambda\delta} J_{\delta.\beta}^{-1} J_{\delta\lambda} - J_{\lambda\tau.\delta} J_{\tau.\beta}^{-1} J_{\tau\lambda}]} \tag{4.36}$$

The unadjusted RS is :

$$RS_{\lambda} = \frac{d_{\lambda}^2}{J_{\lambda}} \tag{4.37}$$

Using these proposed adjusted test statistics, I will address the empirical question at hand, i.e., specification search for growth model. Specifically, I will start with the basic panel model and estimate it by OLS method and then will use the proposed tests to identify the specific sources of departures. Before embarking on data analysis, I discuss some further elegant and useful features of specification search.

4.4.4 Analysis of Misspecification

Earlier I demonstrated that the"time and panel" parameters $(\gamma, \sigma^2_\mu, \rho)$ are orthogonal to the "spatial" parameters (δ, λ, τ) in the sense of testing. Thus the joint RS statistic for $H^a_o : \gamma = \sigma^2_\mu = \rho = \delta = \tau = \lambda = 0$, RS_J decomposes naturally into two orthogonal components:

$$RS_J = RS_{\gamma\sigma^2_\mu\rho} + RS_{\delta\tau\lambda}. \tag{4.38}$$

As I noted there is no further orthogonality among (within) time and panel parameters $(\gamma, \sigma^2_\mu, \rho)$, and spatial parameters (δ, τ, λ). Thus one needs to use the adjusted tests to decompose $RS_{\gamma\sigma^2_\mu\rho}$ and $RS_{\delta\tau\lambda}$ further. From the expressions of different test statistics it follows that

$$RS_{\gamma\sigma^2_\mu\rho} = RS^*_{\gamma|\sigma^2_\mu\rho} + RS_{\sigma^2_\mu\rho} = RS^*_{\gamma|\sigma^2_\mu\rho} + RS^*_{\sigma^2_\mu|\rho} + RS_\rho = RS^*_{\gamma|\sigma^2_\mu\rho} + RS^*_{\rho|\sigma^2_\mu} + RS_{\sigma^2_\mu} \tag{4.39}$$

where $RS^*_{\gamma|\sigma^2_\mu\rho}$ is the adjusted test derived in equation (26), $RS^*_{\sigma^2_\mu|\rho}(RS^*_{\rho|\sigma^2_\mu})$ is the adjusted test statistics for σ^2_μ (ρ) after taking care of the parameter ρ (σ^2_μ). Moreover, the analytical form of $RS^*_{\sigma^2_\mu|\rho}$ and $RS^*_{\rho|\sigma^2_\mu}$ are same as derived in Sen and Bera (2011) under static panel spatial model framework.

Alternatively, one can also write:

$$RS_{\gamma\sigma^2_\mu\rho} = RS^*_{\sigma^2_\mu|\gamma\rho} + RS_{\gamma\rho} = RS^*_{\sigma^2_\mu|\gamma\rho} + RS^*_{\gamma|\rho} + RS_\rho = RS^*_{\sigma^2_\mu|\gamma\rho} + RS\rho|\gamma^* + RS_\gamma \tag{4.40}$$

or,

$$RS_{\gamma\sigma^2_\mu\rho} = RS^*_{\rho|\gamma\sigma^2_\mu} + RS_{\gamma\sigma^2_\mu} = RS^*_{\rho|\gamma\sigma^2_\mu} + RS^*_{\gamma|\sigma^2_\mu} + RS_{\sigma^2_\mu} = RS^*_{\rho|\gamma\sigma^2_\mu} + RS^*_{\sigma^2_\mu|\gamma} + RS_\gamma \tag{4.41}$$

Most computer software reports joint and unadjusted (one-directional) RS tests. Above decomposition suggest that one can obtain all the adjusted test without any extra computation. Similar decomposition will hold for the adjusted RS test statistics for the spatial parameters (δ, τ and λ).

$$RS_{\delta\tau\lambda} = RS^*_{\delta|\tau\lambda} + RS_{\tau\lambda} = RS^*_{\delta|\tau\lambda} + RS^*_{\tau|\lambda} + RS_{\lambda} = RS^*_{\delta|\tau\lambda} + RS^*_{\lambda|\tau} + RS_{\tau} \tag{4.42}$$

Therefore, the proposed adjusted test statistics aid the researcher in model specification with minimum estimation, and its elegant additive property give the researcher a wide flexibility in selecting a model framework.

4.5 Specification Search for Growth Model

4.5.1 Data

The data is from Penn World Tables (PWT, version 6.1), which contain information on real income, investment and population (among many other variables) for a large number of countries. In this paper, I use a sample of 91 countries over the period of 1961 - 1995. These countries are those of MRW (1992) non-oil sample which has been used extensively by other researchers for empirical work on growth convergence.

The dependent variable is real income per worker is measured by real GDP computed by chain method, divided by number of workers. I computed the number of workers following Caselli (2005): $RGDPCH \times POP/RGDPW$, where RGDPCH is real GDP per capita computed by chain method, RGDPW is real GDP per worker and POP is population. The independent variables are same as in MRW (1992). They are n, which measures the average growth of the working-age population (ages 15 to 64), the savings rate s is measured as the average share of gross investment in GDP.

There are many ways one can specify the weight matrix W, for example, geographic distance, k-neighborhood matrix, contiguous neighborhood matrix, economic distance, etc.

(See Conley and Topa (2002), Eaton and Kortun (1996), Klenow and Rodriguez-Clare (2005) and Ertur and Koch (2007).) I consider three different specification of weight matrix, mainly to check the sensitivity of the test result to different definition of spatial connectedness. Here, W_1, W_2 and W_3, where W_1 are defined. The elements of $W_1 are\ w1_{ij} = \frac{d_{ij}^{-2}}{\sum_j d_{ij}^{-2}}$, such that $d_{ii} = 0$ and d_{ij} is the euclidean distance between country's capital. Other two matrices, W_2 and W_3, are based on k-nearest neighbors, with $k = 8$ and 20 respectively, nearness being measured in terms of the geographic distance.

4.5.2 Specification Search

First I present some basic features of the income distribution of the 91 non-oil countries over the 35- year period 1961 - 1995, in Figure 4.1. Figure 4.1 for four groupings of cross sectional averages of per capita real income, where the groups are selected based on initial income of these countries in 1961. [The details of each group is provided in the appendix]. Averages, maximum and minimum are shown across the four panels. Panel A, B, C, D are respectively, for the poorest, middle, rich and richest income groups. A, B and C each is based on 24 countries, and Panel D represents 19 countries. The trajectories in Figure 4.1 provide some idea of the variability in the actual growth trajectories over time within these groupings. It also indicate that some members of each group have substantial prospects of moving into higher income groups over the 35 year period. However, assuming homogeneity of technological progress and speed of convergence (which is usually assumed for β- convergence in cross-sectional studies) rules out this possibility. It is only when one allows for heterogeneity in technological progress over time and over cross-sectional units (while at the same time requiring that the growth rate of technological progress converge to a common constant over time to ensure convergence), then the realistic patterns as shown in Figure 4.1 can emerge (for detail, see Philips and Sul (2003)). Using the proposed specification tests our objective is now to identify a model that captures such essential features of the data.

Table 4.1 reports the joint RS test RS_J for the null $H_o^a : \gamma = \sigma_\mu^2 = \rho = \delta = \tau = \lambda = 0$, joint test for time dynamics (γ), random effect (σ_μ^2) and serial correlation (ρ) $RS_{\gamma\sigma_\mu^2\rho}$, joint RS test for space recursive (δ), spatial lag (τ) and spatial error lag (λ) $RS_{\delta\tau\lambda}$ and Table 4.2 reports the unadjusted single-directional RS tests for each of the six parameters, and the proposed adjusted RS test statistics (noted by '*') for all the parameters. Except $RS_J, RS_{\gamma\sigma_\mu^2\rho}$ and $RS_{\delta\tau\lambda}$, each of the test statistics follow χ_1^2 distribution asymptotically. $RS_J \sim \chi_6^2$, and $RS_{\gamma\sigma_\mu^2\rho}$ and $RS_{\delta\tau\lambda} \sim \chi_3^2$. Each of these tables report the test statistics for same model under three different specification of the weight matrix W. Recall, W_1 uses the geographical distance between the capital of the countries, W_2 and W_3 use eight and twenty nearest neighbors, respectively.

No matter which W is chosen, RS_J is always highly significant at any significance level. Given the orthogonality between spatial and panel-time parameters, as given in the additivity result in equation (4.38), we can conduct the joint tests $RS_{\gamma\sigma_\mu^2\rho}$ and $RS_{\delta\tau\lambda}$. These joint tests are also highly significant after comparing with χ_3^2 critical points at any significance level; however they are not informative about the specific direction(s) of the misspecification(s). Thus based on the results for the unadjusted tests, it would appear that the features like time dynamics, serial correlation of error, random effects, spatial dependencies are the features of this dataset and therefore should be added to the basic model (joint null). However, as we discussed earlier, the inference based on the unadjusted tests can be highly misleading as they fail to take into account the possible presence of other parameters and their interdependencies.

Significance of each parameter can only be evaluated correctly by considering the modified tests. Out of all the six adjusted test statistics, only four, namely, $RS_\gamma^*, RS_{\sigma_\mu^2}^*, RS_\rho^*$ and RS_λ^* are significant, irrespective of the choice of weight matrix W. It is interesting to note the difference in values of the test statistics; the adjusted test statistics are much lower than their unadjusted counterparts. The striking differences in values can be noted for RS_δ, RS_τ with their adjusted counterparts, i.e., RS_δ^* and RS_τ^* respectively. The value of RS_δ^* falls below the critical point of χ_1^2 at any significance level, after it takes into account the possible presence of τ and λ. Similarly spatial lag dependence (τ) parameter loses its significance as the test statistic drops from 7.96 to 2.42 after adjustment. Viewing the test statistics as a measure of degree of misspecification, we find that out of all the departures from the joint null, most of the misspecification is attributed to σ_μ^2 ($\frac{59.17}{231.91} = 26\%$) and time dynamics γ ($\frac{48.11}{231.91} = 21\%$) followed by spatial error dependence (which captures the indirect cross-sectional dependence among these countries). This holds true with W_3; for W_2, however, when the weight matrix is relatively sparse. Thus for W_2, the misspecification due to spatial dependence is relatively low. This may be due to the fact that W_2 is sparse and thus dilute the degree of spatial dependencies. However, no matter what form of W is chosen, misspecification due to time dynamics, serial correlation and random effects are always strong. Thus, when the full sample of annual data is used for 91 non-oil countries for the period 1961 - 1995, the relevant growth regression [2] would be

$$y_{it} = \gamma y_{it-1} + \beta_1 x_{it} + \mu_i + \epsilon_{it} \tag{4.43}$$

$$\epsilon_{it} = \lambda \sum_{j \neq i} w_{ij} \epsilon_{jt} + v_{it} \tag{4.44}$$

$$v_{it} = \rho v_{it-1} + e_{it} \tag{4.45}$$

Note that this specification, interestingly, also supports kind of trajectories Figure 4.1 demonstrates as it can take into account the heterogeneity in technological progress across countries and across time, captured through the spatial dependence (which can account for the technology transfer between countries) and the serial correlation (which captures the differences across time).[3].

MRW (1992) considers a cross-sectional regression model for the non-oil sample with s and n as explanatory variables and finds the rate of conditional convergence be to be very low, 0.00606(0.001) (speed of convergence $= -(1-\hat{\alpha})(n+g+\xi)$, where $\hat{\alpha}$ is the estimated share of physical capital, n is working population, g is growth rate of the country, ξ is depreciation of capital), implying a half-life of 114 years, which is indeed very long. Islam (1995) uses fixed effect dynamic panel data model and allows for the unobserved technological diffusion through the fixed effects term, and estimates the rate of conditional convergence to

[2]Concerns may be raised regarding the estimation of such model specially when the time dynamics is present along with the random effect, as it is generally believed that the inclusion of lagged dependent variables in a panel model necessarily renders random-effect estimators inconsistent. However, it has been shown in Ashley (2010) that if the variables X are strictly exogenous then the lagged value of the quasi-differenced dependent variable is uncorrelated with the quasi-differenced model error, and thus usual random effect estimator would provide consistent estimators of the parameters.

[3]Only assumption required for convergence here is that over time (technically $t \to \infty$) the technological differences between countries i and j should go to zero.

be 0.0434 (0.007) (speed of convergence measured as $\frac{1}{\Delta}\ln\hat{\gamma}$, where Δ is the time difference between two consecutive periods and $\hat{\gamma}$ is the estimate of time dynamics parameter in a fixed effect dynamic panel data model) . It should be noted here, that Islam (1995) used minimum distance (MD) estimator to estimate his model using similar data as MRW(1992). According to Islam (1995), the panel estimate of the convergence rate increases 7.2 times (relative to its OLS estimates that ignores technological differences, as in MRW (1992)) in the non-oil sample, thus concluding that for these countries the half-life is 16 years approximately. Lee et al. (1997) estimate the rate of convergence to be 0.1845 for the same sample of countries by allowing the growth rate g to differ across countries and also for possible serial correlation of error. Ertur and Koch (2007) consider the growth convergence model allowing for regional knowledge and technology spillover effects through spatial dependence, and estimate the rate of convergence to be 0.012 (0.00), with half life around 59 years (speed of convergence $= \frac{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}(n_j+g+\xi)}{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}} - \sum_{j=1}^{N} u_{ij}\frac{1}{\Theta_j}(n_j+g+\xi)$, where u_{ij} is a function of estimates of capital share, spatial dependence and elements of W matrix, Φ_j and Θ_j are the rate of convergence of capital and income, respectively, to the steady state, of country j. For details see, Ertur and Koch (2007).) Thus, the speed of convergence and the implied half-life clearly depends on the model framework. For same dataset, under different frameworks, researchers got widely varying rates of convergence and widely varying half-life estimates (number of years needed for conditional convergence) corresponding to each of the respective convergence estimates. This is a surprising fact and hasn't been considered so far by any research papers. So it is obvious that these models could not capture all the salient feature of the underlying data and thats why the estimate of growth convergence from these models can result in potentially misleading policy implications. The proposed tests can aid the researcher in tackling this difficult task- i.e., to understand the DGP with minimum estimation a priori. As I discussed, given the annual data for the sample of 91 countries over the period 1961 - 1995, the most appropriate model specification is given by equations (3.43) - (3.45).

In Table 4.3 and 4.4, I have considered specification search for growth models under different time frames. I consider the test statistics when the time span is five - year intervals. Thus considering the period 1961 - 1995, I have seven data (time) points for each country: 1965, 1970, 1975, 1980, 1985, 1990, 1995. The variables are averages over five -year time intervals. Islam (1995) pointed out that "yearly time span are too short to be appropriate for growth convergence. Short-term disturbances may loom large in such brief time spans." Following his work, many researchers considered this data setup so that the growth convergence estimates are less influenced by business cycle fluctuations and less likely to be serially correlated than they would be in a yearly data setup. I also divide the data in two subsamples to investigate if the the model specification search is robust. Therefore, RS^*_{S1}, RS_{S1}, RS^*_{S2} and RS_{S2} are respectively the adjusted and unadjusted test statistics for the annual data for 91 countries for the subsamples 1961 - 1980 (referred as S1) and 1981 - 1995 (referred as S2) respectively.

It is evident from Table 4.3 that all the joint tests are significant irrespective of the time span of the data. From column 2 of Table 4.4, it is evident that only the heterogeneity (σ^2_μ) and time dynamics (γ) are most prominent features when 5 year time interval is chosen. RS^*_λ is significant at 5% level. Interestingly although the unadjusted test for ρ is significant, but adjusted one is no longer so. Thus when the data setup is based on 5-year time interval, the relevant growth equation is very similar to Islam (1995), i.e., dynamic panel data framework augmented for cross-sectional dependence also. Indeed 5 -year time span removes effect of serial correlation of errors.

Column 4 - Column 7 of Table 4.4 indicate that the relevant feature of growth regression are similar to equation (4.43) - (4.45). Although the relative values of the test statistics are different, but the inference remains same. This supports the robustness of the results using the proposed test statistics.

To summarize, in this section I use my proposed test statistics from Section 4.4, for the proper specification search of growth model. As explained in Section 4.4.1, I need to estimate *only* the simple panel model in order to apply the proposed tests. Thus, no complex estimation is necessary. I show how one can unravel the salient features of the underlying DGP using the proposed tests. In particular, I show that for the given dataset of 91 non-oil samples from Penn World table, the most relevant features are heterogeneity, time dynamics and indirect cross-sectional dependence. Thus a researcher analyzing the growth behavior of these 91 countries should take care of these departures in his/her model; otherwise the model would be misspecified which would lead to wrong policy implication. For example, as I have shown here, if one assumes a cross-sectional regression model, then the growth convergence rate is very low, implying the half-life to be 114 years. Again assuming a fixed effect dynamic panel model will yield a much higher rate of convergence for the same dataset, implying half-life to be as short as 16 years. It is evident that the convergence rate of income vary wildly, even when same dataset is used. Thus one should consider a proper specification search before directly going into model implication and policy analysis. Tables 4.1 and 4.2 illustrate this important fact and Tables 4.3 and 4.4 demonstrate the robustness of the proposed test result.

In the next section I demonstrate that though the suggested tests are valid only for large samples and local misspecification, they perform quite well in finite samples.

4.6 Monte Carlo Results

The proposed tests are valid only asymptotically. As in our empirical application, in the real world data will be limited. Therefore, we need to evaluate the performance of the tests under a finite sample scenario. The data for Monte Carlo study were generated based on the model (4.1) - (4.4).

$$y_{it} = \gamma y_{it-1} + \tau \sum_{j=1}^{N} m_{ij} y_{jt} + \delta \sum_{j=1}^{N} m_{ij} y_{jt-1} + X_{it}\beta + u_{it}$$

$$u_{it} = \mu_i + \epsilon_{it}$$

$$\epsilon_{it} = \lambda \sum_{j=1}^{N} w_{ij} \epsilon_{jt} + v_{it}$$

$$v_{it} = \rho v_{it-1} + e_{it}, \text{where } e_{it} \sim IIDN(0, \sigma_e^2)$$

We set $\alpha = 5$ and $\beta = 0.5$. The independent variable X_{it} is generated using:

$$X_{it} = 0.4 X_{it-1} + \varphi_{it}$$

where $\varphi \sim Unif[-0.5, 0.5]$ and $X_{i0} = 5 + 10\varphi_{i0}$. For weight matrix W, I consider rook design. I fixed $\sigma_\mu^2 + \sigma_e^2 = 20$ and let $\eta = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_e^2}$. Values of all the six parameters $\gamma, \sigma_\mu^2, \rho, \delta, \tau$ and λ are varied over a range from 0 to 0.5 . I have considered two different pairs for (N, T) namely $(25, 10), (49, 20)$. For lack of space I report the results for $(25, 10)$. The results for $(49, 20)$ are quite comparable to the reported ones and are available on request. Each Monte Carlo experiment is consist of generating 1000 samples for each different parameter settings. Thus the maximum standard error of the estimates of the size and power would be $\sqrt{\frac{(0.5(1-0.5)}{1000}} = 0.015$. The parameters were estimated using OLS, and fifteen test statistics, namely $RS_J, RS_{\gamma\sigma_\mu^2\rho}, RS_{\delta\tau\lambda}, RS_\gamma^*, RS_\gamma, RS_{\sigma_\mu^2}^*, RS_{\sigma_\mu^2}, RS_\rho^*, RS_\rho, RS_\delta^*, RS_\delta, RS_\tau^*,$ RS_τ, RS_λ^* and RS_λ were computed. As discussed earlier, in practice it is not necessary to compute all these statistics ; I do it here only for comparative evaluation. The results are based on the nominal size of 0.05.

As noted in Section 4.4, the parameters $(\gamma, \sigma_\mu^2, \rho)$ are orthogonal to (δ, τ, λ) as far as testing are concerned. So I report the key results in two tables. Table 3.5 reports the size and power of $RS_\gamma^*, RS_\gamma, RS_{\sigma_\mu^2}^*, RS_{\sigma_\mu^2}, RS_\rho^*, RS_\rho$ and Table 6 reports the size and power of $RS_\delta^*, RS_\delta, RS_\tau^*, RS_\tau, RS_\lambda^*$ and RS_λ. Results $RS_J, RS_{\gamma\sigma_\mu^2\rho}$ and $RS_{\delta\tau\lambda}$ are not reported for lack of space. However, each of them achieves nominal size under joint null, and expected power properties.

In Table 4.5, I vary the parameters (γ, η, ρ) from 0 to 0.5, one and two at a time, keeping the spatial parameters zero. RS_γ^* is size robust while RS_γ performs badly when $\gamma = 0$ **and** when either or both $\eta \neq 0$ and $\rho \neq 0$. For example, when $\gamma = 0, \eta = 0, \rho = 0.3$, rejection probability of RS_γ^* is 0.048 and that for RS_γ is 0.888. Similarly, when $\gamma = 0.3, \eta = 0, \rho = 0$ then rejection probability of $RS_{\sigma_\mu^2}^*$ is 0.054 and that of $RS_{\sigma_\mu^2}$ is 0.231. When $\gamma = 0.4, \eta = 0, \rho = 0$ the rejection probability of RS_ρ^* is 0.040 and that of RS_ρ is 0.166. Thus as expected by construction, the adjusted test statistics are size-robust under (local and in some cases even global) misspecification while their unadjusted counter parts are not. However, there is slight loss in power for these adjusted test statistics compared to the unadjusted ones, when adjustments are made even when there is no misspecification. This loss in power however reduces as the parameter values deviates further from the null. As discussed in Sen and Bera (2011), this loss in power can be regarded as the premium one pays for the validity of the adjusted test under local misspecification, i.e., the cost of robustness.

In Table 4.6, I vary the spatial parameters (δ, τ, λ) from 0 to 0.5, one and two at a time, keeping other parameters at zero. It is evident from Table 3.6 that RS^*_δ, RS^*_τ and RS^*_λ are more size-robust than RS_δ, RS_τ and RS_λ respectively. For instance, when $\delta = 0, \tau = 0.4, \lambda = 0.4$, the rejection probability of RS^*_δ is 0.049, whereas for RS_δ it is 0.997. Again when $\delta = 0.4, \tau = 0, \lambda = 0$, the rejection probability for RS^*_λ is 0.038 and that for RS_λ is 0.99. Further Monte Carlo results on $RS^*_\gamma, RS_\gamma, RS^*_{\sigma^2_\mu}, RS_{\sigma^2_\mu}, RS^*_\rho, RS_\rho, RS^*_\delta, RS_\delta, RS^*_\tau, RS_\tau, RS^*_\lambda$ and RS_λ are reported in the Appendix B.

4.7 Conclusion

The growth convergence debate has always occupied a central stage in economics. This is mainly because of the existence of the variety of issues regarding such models, like different forms of convergence, estimation techniques, data, variables, sample and so on. In this paper I address one specific concern, i.e., what is the most appropriate model given the data. Thus the contributions of this paper are twofold. *Firstly*, this paper develops adjusted RS test statistics, which are robust under local misspecification in a dynamic panel model allowing for cross-sectional dependence. *Secondly*, using the proposed tests I address the issue of growth model specification objectively.

To achieve these objectives, robust RS tests for time dynamics, random effect, serial correlation of errors, space-time dynamics and spatial dependence are proposed using Bera and Yoon (1993) test principle. These six adjusted tests are robust under "all" possible misspecification. This robustness is achieved without any estimation of the nuisance parameters. For example, the proposed adjusted RS test for time dynamics is made robust to the presence of random effect, serial correlation of errors, space-time dynamics and spatial dependence. I take care of these possible presence of nuisance parameters using their respective Fisher-Rao score functions. Thus, there is no need to estimate the nuisance parameters as usually it is done for conditional LM and LR tests. The proposed (robust) tests are simple to compute and interpret as they are essentially based only on OLS residuals and score functions. In addition, due to an attractive additive property, the robust tests require very little extra computation. Thus one can compute these robust tests for each parameter from the standard RS tests (joint and marginal). Due to this simplicity in terms of computation, the researchers can identify specific direction(s) to reformulate the basic growth model quite easily.

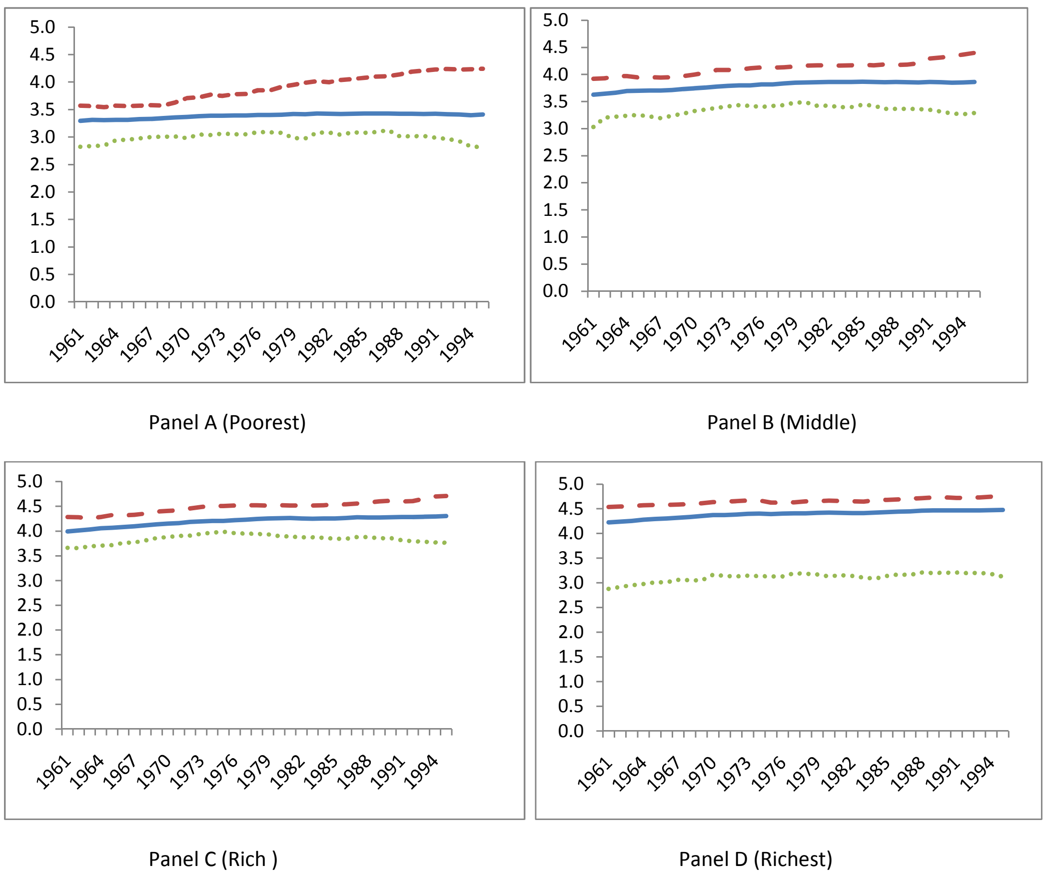
In the empirical application, using these tests, I find that most of the misspecification is attributed to heterogeneity (random effects), dynamic time effects and indirect cross-sectional dependence, irrespective of the specification of weight matrix and time span of the sample. In addition, I demonstrate how the exact nature of dependencies changes the growth model specification for different time framework. Different researchers have derived *widely different convergence rates* for the same dataset, as they considered either only cross-sectional, spatial, panel, or dynamic panel models. It is quite possible that those models cannot capture all the salient feature of the data. Using a model framework which combines all these piece-wise models considered so far in the literature, I conduct the growth model specification search using the proposed test statistics developed for the dynamic panel model with cross-sectional dependence. One should note that the proposed tests are general and can be used for many other specification search of econometric models, for example, hedonic price models, unemployment models. Lastly, through simulation study I demonstrate that the proposed tests, are not only theoretically and asymptotically valid, but can also be used in finite samples exercises where availability of data is often limited.

4.8 Tables and Figures

Table 4.1: Specification Search using Full Sample of 91 countries over 1961 - 1995.

Test Statistics	Specification with W_1	Specification with W_2	Specification with W_3
RS_J	231.91	177.16	217.16
$RS_{\gamma\sigma_\mu^2\rho}$	188.19	141.19	179.19
$RS_{\delta\tau\lambda}$	43.72	35.97	37.97

Figure 4.1: Real per-capita income trajectories of 91 countries: 1961 - 1995



Average (blue ——), Minimum (green......) and Maximum (red - - - -)

Table 4.2: Specification Search using Full Sample of 91 countries over 1961 - 1995.

Parameters	RS_{W_1}	$RS^*_{W_1}$	RS_{W_2}	$RS^*_{W_2}$	RS_{W_3}	$RS^*_{W_3}$
Time-Dynamics - γ	81.12***	48.11***	60.09***	37.51***	79.51***	43.75***
Heterogeneity - σ^2_μ	92.45***	59.17***	79.12***	43.68***	89.03***	54.84***
Serial Correlation - ρ	32.31***	15.79***	24.69***	9.94***	27.23***	12.55***
Space-Time Dyn - δ	8.51***	**1.11**	6.42**	**0.19**	7.50***	**0.12**
Spatial Lag - τ	7.96***	**2.42**	5.91**	**1.74**	6.96***	**1.82**
Spatial Error - λ	41.92***	40.68***	33.31***	31.17***	36.14***	35.57***

Note: * indicates significant at 10%, ** indicates significant at 5% and *** indicates significant at 1%.

Table 4.3: Specification search using different time specification.

Test Statistics	5-year time interval	Subsample: 1961 - 1980	Subsample: 1981 - 1995
RS_J	87.41	170.82	154.86
$RS_{\gamma\sigma^2_\mu\rho}$	70.13	145.14	131.47
$RS_{\delta\tau\lambda}$	17.28	25.68	23.39

Table 4.4: Specification search using different time specification.

Parameters	$RS_{5-years}$	$RS^*_{5-years}$	RS_{S1}	RS^*_{S1}	RS_{S2}	RS^*_{S2}
Time-Dynamics - γ	18.50***	10.96***	72.29***	37.11***	69.51***	33.75***
Heterogeneity - σ^2_μ	25.27***	14.07***	89.31***	41.11***	80.23***	39.17***
Serial Correlation - ρ	14.01***	**2.80**	21.69***	6.94***	24.23***	7.55***
Space-Time Dyn - δ	8.05***	**1.31**	5.33**	**1.56**	7.68***	**1.87**
Spatial Lag - τ	**1.01**	**0.009**	4.79**	**1.19**	6.45***	**1.43**
Spatial Error - λ	15.16***	14.22**	22.11***	20.17***	21.14***	19.81***

Note: * indicates significant at 10%, ** indicates significant at 5% and *** indicates significant at 1%.
S1: Subsample: 1961 - 1980, S2: Subsample: 1980 - 1995.

Table 4.5: Estimated Rejection Probabilities with $\delta = \tau = \lambda = 0$. Sample size: $N = 25, T = 10$

γ	η	ρ	RS^*_γ	RS_γ	$RS^*_{\sigma^2_\mu}$	$RS_{\sigma^2_\mu}$	RS^*_ρ	RS_ρ
0	0	0	**0.055**	**0.057**	**0.058**	**0.051**	**0.058**	**0.054**
0.1	0	0	0.058	0.497	0.069	0.091	0.115	0.198
0.2	0	0	0.060	0.769	0.054	0.107	0.105	0.291
0.3	0	0	0.048	0.888	0.035	0.131	0.267	0.316
0.4	0	0	0.049	1.000	0.066	0.108	0.277	0.460
0.5	0	0	0.044	1.000	0.041	0.344	0.314	0.698
0	0.4	0	0.054	0.682	0.119	0.099	0.034	0.110
0.1	0.4	0	0.068	0.798	0.155	0.112	0.296	0.332
0.2	0.4	0	0.084	0.894	0.151	0.202	0.287	0.413
0.3	0.4	0	0.078	0.959	0.238	0.346	0.375	0.534
0.4	0.4	0	0.081	1.000	0.324	0.425	0.451	0.611
0.5	0.4	0	0.091	1.000	0.211	0.492	0.524	0.723
0	0.1	0	0.065	0.753	0.161	0.212	0.045	0.111
0	0.2	0	0.065	0.861	0.166	0.202	0.032	0.203
0	0.3	0	0.072	0.952	0.191	0.346	0.035	0.298
0	0.4	0	0.097	0.955	0.206	0.425	0.044	0.229
0	0.5	0	0.069	0.965	0.263	0.492	0.054	0.398
0.4	0	0	0.014	1.000	0.043	0.109	0.230	0.157
0.4	0.1	0	0.015	1.000	0.113	0.129	0.239	0.264
0.4	0.2	0	0.019	1.000	0.217	0.258	0.329	0.319
0.4	0.3	0	0.029	1.000	0.217	0.254	0.354	0.401
0.4	0.4	0	0.029	1.000	0.323	0.432	0.448	0.503
0.4	0.5	0	0.045	1.000	0.343	0.521	0.570	0.512
0	0	0.1	0.153	0.745	0.078	0.071	0.049	0.109
0	0	0.2	0.127	0.865	0.067	0.207	0.044	0.111
0	0	0.3	0.231	1.000	0.054	0.231	0.046	0.209
0	0	0.4	0.226	1.000	0.045	0.208	0.040	0.166
0	0	0.5	0.334	1.000	0.066	0.344	0.043	0.254
0.4	0	0	0.044	1.000	0.053	0.111	0.221	0.188
0.4	0	0.1	0.302	1.000	0.091	0.201	0.210	0.582
0.4	0	0.2	0.401	1.000	0.061	0.294	0.311	0.889
0.4	0	0.3	0.504	1.000	0.051	0.363	0.333	0.965
0.4	0	0.4	0.534	1.000	0.076	0.388	0.355	0.994
0.4	0	0.5	0.621	1.000	0.056	0.424	0.368	0.996

Table 4.6: Estimated Rejection Probabilities with $\gamma = \sigma_\mu^2 = \rho = 0$. Sample size: $N = 25, T = 10$

δ	τ	λ	RS_δ^*	RS_δ	RS_τ^*	RS_τ	RS_λ^*	RS_λ
0	0	0	**0.049**	**0.056**	**0.053**	**0.054**	**0.047**	**0.061**
0	0	0.1	0.047	0.694	0.161	0.783	0.052	0.953
0	0	0.2	0.045	0.786	0.133	0.950	0.157	0.995
0	0	0.3	0.051	0.964	0.255	0.992	0.372	1.000
0	0	0.4	0.061	0.933	0.235	1.000	0.635	1.000
0	0	0.5	0.039	0.863	0.356	1.000	0.861	1.000
0	0.4	0	0.051	1.000	0.965	1.000	0.031	1.000
0	0.4	0.1	0.038	1.000	0.990	1.000	0.044	1.000
0	0.4	0.2	0.039	1.000	0.995	1.000	0.128	1.000
0	0.4	0.3	0.047	0.992	0.993	1.000	0.229	1.000
0	0.4	0.4	0.057	0.980	0.982	1.000	0.464	1.000
0	0.4	0.5	0.060	0.943	0.939	1.000	0.780	1.000
0	0.1	0	0.037	0.782	0.284	0.893	0.058	0.759
0	0.2	0	0.031	0.876	0.867	0.992	0.070	0.996
0	0.3	0	0.051	0.934	0.934	1.000	0.067	0.999
0	0.4	0	0.051	1.000	0.961	1.000	0.065	1.000
0	0.5	0	0.041	1.000	0.979	1.000	0.078	1.000
0	0.1	0.4	0.048	0.951	0.950	1.000	0.648	1.000
0	0.2	0.4	0.053	0.968	0.964	1.000	0.640	1.000
0	0.3	0.4	0.037	0.974	0.971	1.000	0.487	1.000
0	0.4	0.4	0.049	0.994	0.985	1.000	0.479	1.000
0	0.5	0.4	0.056	0.986	0.990	1.000	0.468	1.000
0.1	0	0	0.238	0.798	0.114	0.712	0.055	0.789
0.2	0	0	0.333	0.894	0.267	0.871	0.068	0.849
0.3	0	0	0.357	0.976	0.334	0.967	0.054	0.977
0.4	0	0	0.459	1.000	0.261	0.989	0.038	0.990
0.5	0	0	0.600	1.000	0.379	0.996	0.090	0.993
0	0	0.4	0.049	0.947	0.143	0.999	0.674	1.000
0.1	0	0.4	0.368	0.991	0.189	1.000	0.445	0.999
0.2	0	0.4	0.410	1.000	0.292	1.000	0.301	1.000
0.3	0	0.4	0.505	1.000	0.395	1.000	0.295	1.000
0.4	0	0.4	0.601	1.000	0.391	1.000	0.376	1.000
0.5	0	0.4	0.632	1.000	0.492	1.000	0.448	1.000

APPENDIX A

TECHNICAL APPENDIX FOR CHAPTER 3

A.1 Background

We consider the following panel spatial model which is the combination of all the different piecewise framework which has been discussed in Chapter 3.

$$y_{it} = \tau \sum_{j=1}^{N} m_{ij} y_{jt} + X_{it}\beta + u_{it} \tag{A.1}$$

$$u_{it} = \mu_i + \epsilon_{it} \tag{A.2}$$

$$\epsilon_{it} = \lambda \sum_{j=1}^{N} w_{ij} \epsilon jt + v_{it} \tag{A.3}$$

$$v_{it} = \rho v_{it-1} + e_{it}, \text{where } e_{it} \sim IIDN(0, \sigma_e^2) \tag{A.4}$$

for $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$. Here y_{it} is the observation for i^{th} individual/observation at t^{th} time, X_{it} denotes the observations on non-stochastic regressors and u_{it} is the regression disturbance. Spatial dependence is captured by the weigh matrices $M = (m_{ij})$ and $W = (w_{ij})$. In this framework, I have considered spatial lag dependence (τ), spatial error dependence (λ), serial correlation in error (ρ) and individual effect (μ_i). I consider random effect model here, i.e., $\mu_i \sim IID(0, \sigma_\mu)$. W and M are row-standardized weight matrices whose diagonal elements are zero, such that $(I - \tau M)$ and $(I - \lambda W)$ are non-singular, where I is an identity matrix of dimension N. I assume that the model satisfies the regularity conditions given in Lee and Yu (2010).

In matrix form, the equations (A.1) - (A.4) can be rewritten as

$$y = \tau(I_T \otimes M)y + X\beta + u \tag{A.5}$$

where y is f dimension $NT \times 1, X$ is $NT \times K, \beta$ is $k \times 1$ and u is $NT \times 1$. Here l is the lag operator, X is assumed to be of full column rank and its elements are bounded in absolute value. The disturbance term can be expressed as

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})v \tag{A.6}$$

where $B = (I_N - \lambda W), \iota_T$ is vector of ones of dimension T, I_T is an identity matrix of dimension $T \times T$ and $\otimes$ denotes Kronecker product. Under this setup, the variance-covariance matrix of u is given by

$$\Omega = \sigma_{\mu^2}[J_T \otimes I_N] + [V \otimes (B'B)^{-1}] \tag{A.7}$$

where J_T is a matrix of ones of dimension $T \times T$, and V is the familiar $T \times T$ variance -covariance matrix for AR (1) process in equation (6.4), i.e.,

$$V = E(v'v) = [\frac{1}{1-\rho^2}V_1] \otimes \sigma_e^2 I_N = V_\rho \otimes \sigma_e^2 I_N \tag{A.8}$$

with

$$V_1 = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \vdots & \vdots & \ddots & \vdots & \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 & \end{bmatrix}$$

and $V_\rho = \frac{1}{1-\rho^2}V_1$.

The loglikelihood function of the above model can be written as:

$$L = \frac{-NT}{2} ln 2\pi - \frac{1}{2}\ln|\Omega| + T\ln|A| - \frac{1}{2}[(I_T \otimes A)y - X\beta]'\Omega^{-1}[(I_T \otimes A)y - X\beta] \tag{A.9}$$

where $A = (I_N - \tau M)$ and following Baltagi et al. (2007), I can write

$$\frac{1}{2}\ln|\Omega| = -\frac{N}{2}\ln(1-\rho^2) + \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| + \frac{NT}{2}\ln\sigma_e^2 - (T-1)\ln|B|$$

where $d^2 = \alpha^2 + (T-1)$, $\alpha = \sqrt{\frac{1+\rho}{1-\rho}}$ and $\phi = \frac{\sigma_\mu^2}{\sigma_e^2}$. Thus substituting $\frac{1}{2}\ln|\Omega|$ in L , I obtain

$$L = \frac{-NT}{2} ln 2\pi + \frac{N}{2}\ln(1-\rho^2) - \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| - \frac{NT}{2}\ln\sigma_e^2 + (T-1)\ln|B| + T\ln|A| - \frac{1}{2}[(I_T\otimes A)y - X\beta]'\Omega^{-1}[(I_T\otimes A)y - X\beta] \quad (A.10)$$

A.2 Derivation of Score

Using the log-likelihood function, eq. (A.10), we derive the following scores.

$$\frac{\partial L}{\partial \beta} = X'\Omega^{-1}u \quad (A.11)$$

$$\frac{\partial L}{\partial \sigma_e^2} = -\frac{1}{2}trC^{-1}\frac{(d^2(1-\rho)^2\sigma_\mu^2 I_N}{\sigma_e^4} - \frac{NT}{2\sigma_e^2} - \frac{1}{2}u'(\Omega^{-1}(V_\rho\otimes[(B'B)^{-1}])\Omega^{-1})u \quad (A.12)$$

$$\frac{\partial L}{\partial \sigma_\mu^2} = -\frac{1}{2}trC^{-1}\frac{(d^2(1-\rho)^2 I_N}{\sigma_e^2} + \frac{1}{2}u'\Omega^{-1}(J_T\otimes I_N)\Omega^{-1}u \quad (A.13)$$

$$\frac{\partial L}{\partial \rho} = -\frac{N\rho}{1-\rho^2} + \frac{1}{2}trC^{-1}(\rho + (T-1)(1-\rho)\phi I_N) + \frac{1}{2}u'(\sigma_e^{-2}(\frac{1}{1-\rho^2})^2[2\rho V_1 + (1-\rho^2)F_\rho]\otimes(B'B)^{-1})u \quad (A.14)$$

$$\frac{\partial L}{\partial \tau} = -Ttr(A^{-1}W) + \frac{1}{2}\Omega^{-1}(I_T\otimes W)y \quad (A.15)$$

$$\frac{\partial L}{\partial \lambda} = -(T-1)tr(B^{-1}W) + \frac{1}{2}trC^{-1}[(B'B)^{-1}[B'W + W'B](B'B)^{-1}] - \frac{1}{2}u'\Omega^{-1}(V_\rho\otimes(B'B)^{-1})\Omega^{-1}u \quad (A.16)$$

where $C = (d^2(1-\rho)^2\phi I_N + (B'B)^{-1})$. The score functions evaluated under H_0^a, i.e., restricted MLE of θ_0 with $\tilde{\omega} = (\tilde{\beta}, \tilde{\sigma_e^2})$ are:

$$\frac{\partial L}{\partial \beta} = 0 \quad (A.17)$$

$$\frac{\partial L}{\partial \sigma_e^2} = 0 \tag{A.18}$$

$$\frac{\partial L}{\partial \sigma_\mu^2} = \frac{NT}{2\tilde{\sigma_e^2}}[\frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1] \tag{A.19}$$

$$\frac{\partial L}{\partial \rho} = \frac{NT}{2}[\frac{\tilde{u}'(G \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{A.20}$$

$$\frac{\partial L}{\partial \tau} = \frac{\tilde{u}'[(I_T \otimes W)Y_{NT}]}{\tilde{\sigma_e^2}} \tag{A.21}$$

$$\frac{\partial L}{\partial \lambda} = \frac{NT}{2}[\frac{\tilde{u}'(I_T \otimes (W + W'))\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{A.22}$$

where $\tilde{u} = y - x\tilde{\beta}$ is the OLS residual vector, and $\tilde{\sigma_e^2} = \frac{\tilde{u}'\tilde{u}}{NT}$.

A.3 Derivation of Information Matrix

To derive the information matrix under joint null H_o^a, I need to derive the second order derivatives and take expectation. Differentiating equation (A.11) w.r.t $\beta, \sigma_e^2, \sigma_\mu^2, \rho, \tau$ and λ, we get

$$\frac{\partial^2 L}{\partial \beta \partial \beta'} = -X'\Omega^{-1}X \tag{A.23}$$

$$\frac{\partial^2 L}{\partial \beta \partial \sigma_e^2} = -u'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}X \tag{A.24}$$

$$\frac{\partial^2 L}{\partial \beta \partial \sigma_\mu^2} = -X'\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \tag{A.25}$$

$$\frac{\partial^2 L}{\partial \beta \partial \rho} = -X'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{A.26}$$

$$\frac{\partial^2 L}{\partial \beta \partial \tau} = -X'\Omega^{-1}(I_T \otimes W)Y \tag{A.27}$$

$$\frac{\partial^2 L}{\partial \beta \partial \lambda} = -X'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{A.28}$$

Differentiating equation (A.12) w.r.t σ_μ^2, ρ, τ and λ, we get

$$\frac{\partial^2 L}{\partial\sigma_e^2\partial\sigma_\mu^2} = \frac{1}{2}tr[C^{-1}\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}C^{-1}(B'B)^{-1}] - u\Omega^{-1}(J_T\otimes I_N)\Omega^{-1}(V_\rho\otimes(B'B)^{-1})\Omega^{-1}u \tag{A.29}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial\sigma_e^2\partial\rho} &= -tr[C^{-1}[(\rho+(1-\rho)(T-1))\phi I_N]C^{-1}(B'B)^{-1}]\\ &\quad - u\Omega^{-1}[[2\frac{\rho}{1-\rho^2}^2 V_1 + \frac{1}{1-\rho^2}F_\rho]\otimes(B'B)^{-1}]\Omega^{-1}(V_\rho\otimes(B'B)^{-1})\Omega^{-1}u\\ &\quad + \frac{1}{2}u\Omega^{-1}[[\frac{2\rho}{(1-\rho^2)^2}V_1 + \frac{1}{1-\rho^2}F_\rho]\otimes(B'B)^{-1}]\end{aligned} \tag{A.30}$$

$$\frac{\partial^2 L}{\partial\sigma_e^2\partial\tau} = -((W\otimes I_T)Y_{NT})'\Omega^{-1}(V_\rho\otimes(B'B)^{-1})\Omega^{-1}u \tag{A.31}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial\sigma_e^2\partial\lambda} &= -\frac{1}{2}tr[C^{-1}(B'B)^{-1}(B'W+W'B)(B'B)^{-1}C^{-1}(B'B)^{-1}+\\ &\quad C^{-1}(B'B)^{-1}(B'W+W'B)(B'B)^{-1}] - \frac{T-1}{2}tr[(B'B)^{-1}(B'W+W'B)(B'B)^{-1}]\\ &\quad - u\Omega^{-1}[V_\rho\otimes(B'B)^{-1}[B'W+W'B](B'B)^{-1}]\Omega^{-1}\\ &\quad (V_\rho\otimes(B'B)^{-1})\Omega^{-1}u + \frac{1}{2}u\Omega^{-1}(V_\rho\otimes(B'B)^{-1}(B'W+W'B)(B'B)^{-1})\Omega^{-1}u\end{aligned} \tag{A.32}$$

Differentiating equation (A.13) w.r.t σ_μ^2, ρ, τ and λ, we get

$$\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\sigma_\mu^2} = \frac{1}{2}tr[C^{-1}\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}C^{-1}\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}] - u\Omega^{-1}(J_T\otimes I_N)\Omega^{-1}(J_T\otimes I_N)\Omega^{-1}u \tag{A.33}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\rho} &= \frac{1}{2}tr[\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}[C^{-1}(\rho+(T-1)(1-\rho)\phi I_N C^{-1})] + u'\Omega^{-1}[\sigma_e^2(\frac{1}{1-\rho^2})^2[2\rho V_1\\ &\quad + (1-\rho^2)F_\rho]\otimes(B'B)^{-1}]\Omega^{-1}(J_T\otimes I_N)\Omega^{-1}u\end{aligned} \tag{A.34}$$

$$\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\tau} = -[(W\otimes I_T)Y_{NT}]'\Omega^{-1}(J_T\otimes I_N)\Omega^{-1}u \tag{A.35}$$

$$\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\lambda} = -\frac{1}{2}tr[\frac{d^2(1-\rho)^2}{\sigma_e^2}I_N C^{-1}\frac{d^2(1-\rho)^2}{\sigma_e^2}I_N C^{-1}]$$
$$+ u'\Omega^{-1}[\sigma_e^2(V_\rho \otimes [(B'B)^{-1}[B'W + W'B](B'B)^{-1}]\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \quad (A.36)$$

Differentiating equation (A.14) wrt ρ, τ and λ, we get

$$\frac{\partial^2 L}{\partial\rho\partial\rho} = \frac{-N + N\rho^2}{(1-\rho^2)^2} + \frac{1}{2}tr((2\rho + (T-1)(1-\rho))\phi I_N)[C^{-1}(\rho + (T-1)(1-\rho))\phi I_N C^{-1}]$$
$$+ u\Omega^{-1}[\sigma_e^2(\frac{1}{1-\rho^2})^2[2\rho V_1 + (1-\rho^2)F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}[\sigma_e^2$$
$$(\frac{1}{1-\rho^2})^2[2\rho V_1 + (1-\rho^2)F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}u + u\Omega^{-1}[[\sigma_e^2(\frac{1}{1-\rho^2})^2$$
$$[2V_1 - 2\rho F_\rho] + 4(1-\rho^2)\rho(2\rho V_1 + (1-\rho^2)F_\rho) \otimes (B'B)^{-1}]\Omega^{-1}u \quad (A.37)$$

$$\frac{\partial^2 L}{\partial\rho\partial\tau} = -[(W \otimes I_T)Y_{NT}]\Omega^{-1}[[\frac{2\rho}{(1-\rho^2)^2}V_1 + \frac{1}{1-\rho^2}F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}u \quad (A.38)$$

$$\frac{\partial^2 L}{\partial\rho\partial\lambda} = \frac{1}{2}tr[[(B'B)^{-1}[B'W + W'B](B'B)^{-1}C6{-}1(\rho + (T-1)(1-\rho))\phi I_N C^{-1}]$$
$$+ u\Omega^{-1}[\sigma_e^2(\frac{1}{1-\rho^2})^2[2V_1 - 2\rho F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}[\sigma_e^2(V_\rho \otimes [(B'B)^{-1}[B'W + W'B](B'B)^{-1}]\Omega^{-1} -$$
$$\frac{1}{2}u\Omega^{-1}[\sigma_e^4(\frac{1}{1-\rho^2})^2[2V_1 - 2\rho F_\rho] \otimes [(B'B)^{-1}[B'W + W'B](B'B)^{-1}](B'B)^{-1}]\Omega^{-1}u \quad (A.39)$$

Differentiating equation (A.15) wrt τ and λ, we get

$$\frac{\partial^2 L}{\partial\tau\partial\tau} = -Ttr((A^{-1}W)^2) - Y_{NT}(I_T \otimes W)\Omega^{-1}(I_T \otimes W)Y_{NT} \quad (A.40)$$

$$\frac{\partial^2 L}{\partial\tau\partial\lambda} = -u\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}(I_T \otimes W)u \quad (A.41)$$

Under the joint null $H_o^a : \gamma = \sigma_\mu^2 = \rho = \delta = \tau = \lambda = 0$, the non-zero second-order derivatives are :

$\frac{\partial^2 L}{\partial\beta\partial\beta} = -\frac{X'X}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\beta\partial\tau} = -\frac{X'(W \otimes I_T)X\hat{\beta}}{\hat{\sigma_e^2}}$

$\frac{\partial^2 L}{\partial\sigma_e^2\partial\sigma_e^2} = -\frac{NT}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\sigma_e^2\partial\sigma_\mu^2} = -\frac{NT}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\sigma_\mu^2} = -\frac{NT^2}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\rho} = -\frac{N(T-1)}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\rho\partial\rho} = -N(T-1)$
$\frac{\partial^2 L}{\partial\tau\partial\tau} = -(Ttr(W^2+WW') + \frac{(\hat{\beta}'X'(I_T\otimes W')(I_T\otimes W)X\hat{\beta})}{\hat{\sigma_e^2}})$
$\frac{\partial^2 L}{\partial\tau\partial\lambda} = \frac{\partial^2 L}{\partial\lambda\partial\lambda} = -Ttr(W^2+WW')$.

All the other second derivatives becomes zero under joint null. Thus the information matrix J, under H_o^a is

The information matrix J, under H_o^a is

$$J(\theta_0) = \begin{bmatrix} \frac{X'X}{\tilde{\sigma_e^2}} & 0 & 0 & 0 & 0 & \frac{X'(I_T\otimes W)X\tilde{\beta}}{\tilde{\sigma_e^2}} \\ 0 & \frac{NT}{2\tilde{\sigma_e^4}} & \frac{NT}{2\tilde{\sigma_e^4}} & 0 & 0 & 0 \\ 0 & \frac{NT}{2\tilde{\sigma_e^4}} & \frac{NT}{2\tilde{\sigma_e^4}} & \frac{N(T-1)}{\tilde{\sigma_e^2}} & 0 & 0 \\ 0 & 0 & \frac{N(T-1)}{\tilde{\sigma_e^2}} & N(T-1) & 0 & 0 \\ 0 & 0 & 0 & 0 & Ttr(W^2+WW') & Ttr(W^2+WW') \\ \frac{X'(I_T\otimes W)X\tilde{\beta}}{\tilde{\sigma_e^2}} & 0 & 0 & 0 & Ttr(W^2+WW') & H \end{bmatrix} \tag{A.42}$$

where $J = E(-\frac{1}{NT}\frac{\partial^2 L}{\partial\tilde{\theta}\partial\tilde{\theta}'})$ evaluated at θ_0.

A.4 Derivation of test statistics

$$RS_\psi^* = \frac{1}{N}[d_\psi(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})d_\phi(\tilde{\theta})][J_{\psi.\omega}(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})J_{\phi\psi.\omega}(\tilde{\theta})]^{-1}$$
$$[d_\psi(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})d_\phi(\tilde{\theta})]' \tag{A.43}$$

where $\omega = (\beta', \sigma_e^2)', \psi$ and ϕ are different combinations of the parameters $(\gamma, \sigma_\mu^2, \rho, \delta, \tau, \lambda)$.

I) $H_o^c : \sigma_\mu^2 = 0$ in presence of ρ, τ, λ.

Here $\phi = (\rho, \tau, \lambda)$

$d_\psi = d_{\sigma_{\mu^2}}$

$d_\phi = (d_\rho, d_\tau, d_\lambda)$

$J_{\psi\phi.\omega} = (J_{\sigma_\mu^2.\sigma_e^2}, J_{\sigma_\mu^2 \rho}, 0, 0)$

$$J_{\phi.\omega} = \begin{bmatrix} J_\rho & 0 & 0 \\ 0 & J_{\lambda\tau} & J_\lambda \\ 0 & J_{\tau.\beta} & J_{\tau\lambda} \end{bmatrix}$$

The adjusted RS test statistics is:

$$RS^*_{\sigma_\mu^2} = \frac{[d_{\sigma_\mu^2} - J_{\sigma_\mu^2 \rho} J_\rho^{-1} d_\rho]^2}{J_{\sigma_\mu^2.\sigma_e^2} - J_{\sigma_\mu^2 \rho} J_\rho^{-1} J_{\rho\sigma_\mu^2}} = \frac{NT^2(A-B)^2}{2(T-1)(T-2)} \tag{A.44}$$

where $A = \frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1$ and $B = \frac{\tilde{u}'(G \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}}$.

II) $H_o^d : \rho = 0$ in presence of $\sigma_\mu^2, \tau, \lambda$.

Here $\phi = (\sigma_\mu^2, \tau, \lambda)$.

$d_\psi = d_\rho$

$d_\phi = (d_{\sigma_{\mu^2}}, d_\tau, d_\lambda)$

$J_{\psi\phi.\omega} = (J_{\rho\sigma_\mu^2}, 0, 0)$.

$$J_{\phi.\omega} = \begin{bmatrix} J_{\sigma_\mu^2.\sigma_e^2} & 0 & 0 \\ 0 & J_{\lambda\tau} & J_\lambda \\ 0 & J_{\tau.\beta} & J_{\tau\lambda} \end{bmatrix}$$

The adjusted test statistic is:

$$RS^*_\rho = \frac{[d_\rho - J_{\rho\sigma_\mu^2} J^{-1}_{\sigma_{\mu^2}.\sigma_e^2} d_{\sigma_\mu^2}]^2}{J_\rho - J_{\rho\sigma_\mu^2} J^{-1}_{\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2 \rho}} = \frac{NT^2(B - \frac{2A}{T})^2}{4(T-1)(T-\frac{2}{T})} \tag{A.45}$$

III) $H_o^g : \lambda = 0$ in presence of σ_μ^2, ρ, τ.

Here, $\phi = (\sigma_\mu^2, \rho, \tau)$

$d_\psi = d_\lambda$

$d_\phi = (d_{\sigma_{\mu^2}}, d_\rho, d_\tau)$

$J_{\psi\phi.\omega} = (0, 0, J_{\lambda\tau})$.

$$J_{\phi.\omega} \begin{bmatrix} J_{\sigma_\mu^2.\sigma_e^2} & J_{\sigma_\mu^2\rho} & 0 \\ J_{\rho\sigma_\mu^2} & J_\rho & 0 \\ 0 & 0 & J_{\tau.\beta} \end{bmatrix}$$

The adjusted test statistic is:

$$RS_\lambda^* = \frac{[d_\lambda - J_{\lambda\tau}J_{\tau.\beta}^{-1}d_\tau]^2}{J_\lambda - J_{\lambda\tau}J_{\tau.\beta}^{-1}J_{\tau\lambda}} = \frac{ZZ'}{\tilde{T}[1 - \tilde{T}J_{\tau.\beta}^{-1}]} \tag{A.46}$$

where $Z = \frac{1}{2\tilde{\sigma_e^2}}[\tilde{u}'E\tilde{y} - \tilde{T}J_{\tau.\beta}^{-1}(\tilde{u}(E+E')\tilde{u})]$, $E = (I_T \otimes W)$ and $\tilde{T} = Ttr(W^2 + WW')$.

IV) $H_o^f : \tau = 0$ in presence of $\sigma_\mu^2, \rho, \lambda$.

Here $\phi = (\sigma_\mu^2, \rho, \lambda)$

$d_\psi = d_\tau$

$d_\phi = (d_{\sigma_{\mu^2}}, d_\rho, d_\lambda)$

$J_{\psi\phi.\omega} = (0, 0, J_{\tau\lambda})$.

$$J_{\phi.\omega} \begin{bmatrix} J_{\sigma_\mu^2.\sigma_e^2} & J_{\sigma_\mu^2\rho} & 0 \\ J_{\rho\sigma_\mu^2} & J_\rho & 0 \\ 0 & 0 & J_\lambda \end{bmatrix}$$

The adjusted test statistic is:

$$RS_\tau^* = \frac{[d_\tau - J_{\tau\lambda}J_\lambda^{-1}d_\lambda]^2}{J_{\tau.\beta} - J_{\tau\lambda}J_\lambda^{-1}J_{\lambda\tau}} = \frac{[\frac{1}{2\tilde{\sigma_e^2}}[\tilde{u}'E\tilde{y} - (\tilde{u}(E+E')\tilde{u})]]^2}{J_{\tau.\beta} - \tilde{T}} \tag{A.47}$$

A.5 Derivation of partial covariances

A) $J_{\rho(\lambda\tau).\sigma_\mu^2\gamma} = ?$

Here $\psi = \rho, \phi = (\lambda, \tau), \theta = (\sigma_\mu^2, \gamma)$.

$$J_{\psi\phi.\theta} = J_{\psi\phi} - J_{\phi\theta} J_\theta^{-1} J_{\theta\phi} = 0 - [0 \ \ J_{\rho\sigma_\mu^2} \ \ 0] \begin{bmatrix} J_\beta & 0 & 0 \\ 0 & J_{\sigma_\mu^2} & J_{\sigma_\mu^2\sigma_e^2} \\ 0 & J_{\sigma_e^2\sigma_\mu^2} & J_{\sigma_e^2} \end{bmatrix}^{-1} \begin{bmatrix} 0 & J_{\beta\tau} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

Therefore, $J_{\rho(\lambda\tau).\sigma_\mu^2\gamma} = 0$

B) $J_{\sigma_\mu^2(\lambda\tau).\rho\gamma} =?$

Here $\psi = \sigma_\mu^2, \phi = (\lambda, \tau), \theta = (\rho, \gamma)$.

$$J_{\psi\phi.\theta} = J_{\psi\phi} - J_{\phi\theta} J_\theta^{-1} J_{\theta\phi} = 0 - [0 \ \ J_{\rho\sigma_\mu^2} \ \ J_{\sigma_\mu^2\rho}] \begin{bmatrix} J_\beta & 0 & 0 \\ 0 & J_\rho & 0 \\ 0 & 0 & J_{\sigma_e^2} \end{bmatrix}^{-1} \begin{bmatrix} 0 & J_{\beta\tau} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

Therefore, $J_{\sigma_\mu^2(\lambda\tau).\rho\gamma} = 0$

C) $J_{\lambda(\sigma_\mu^2\rho).\tau\gamma} =?$

Here $\psi = \lambda, \phi = (\sigma_\mu^2, \rho), \theta = (\tau, \gamma)$.

$$J_{\psi\phi.\theta} = J_{\psi\phi} - J_{\phi\theta} J_\theta^{-1} J_{\theta\phi} = 0 - [0 \ \ J_{\lambda\tau} \ \ 0] \begin{bmatrix} J_\beta & J_{\beta\tau} & 0 \\ J_\tau & J_{\tau\beta} & 0 \\ 0 & 0 & J_{\sigma_e^2} \end{bmatrix}^{-1} \begin{bmatrix} 0 & J_{\beta\tau} \\ 0 & 0 \\ J_{\sigma_\mu^2\sigma_e^2} & 0 \end{bmatrix} = 0$$

Therefore, $J_{\lambda(\sigma_\mu^2\rho).\tau\gamma} = 0$

D) $J_{\tau(\sigma_\mu^2\rho).\lambda\gamma} =?$

Here $\psi = \tau, \phi = (\sigma_\mu^2, \rho), \theta = (\lambda, \gamma)$.

$$J_{\psi\phi.\theta} = J_{\psi\phi} - J_{\phi\theta} J_\theta^{-1} J_{\theta\phi} = 0 - [0 \ \ J_{\lambda\tau} \ \ J_{\tau\beta}] \begin{bmatrix} J_\beta & 0 & 0 \\ 0 & J_\lambda & 0 \\ 0 & 0 & J_{\sigma_e^2} \end{bmatrix}^{-1} \begin{bmatrix} 0 & J_{\beta\tau} \\ 0 & 0 \\ J_{\sigma_\mu^2\sigma_e^2} & 0 \end{bmatrix} = 0$$

Therefore, $J_{\tau(\sigma_\mu^2\rho).\lambda\gamma} = 0$

E) $J_{\lambda\tau(\sigma_\mu^2\rho).\gamma} =?$

Here $\psi = \tau, \lambda, \phi = (\sigma_\mu^2, \rho), \theta = (\gamma)$.

$$J_{\psi\phi.\theta} = J_{\psi\phi} - J_{\phi\theta} J_\theta^{-1} J_{\theta\phi} = 0 - \begin{bmatrix} J_{\tau\beta} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} J_\beta & 0 \\ 0 & J_{\sigma_e^2} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ J_{\sigma_\mu^2 \sigma_e^2} & 0 \end{bmatrix} = 0$$

Therefore, $J_{\lambda\tau(\sigma_\mu^2 \rho).\gamma} = 0$

APPENDIX B

TECHNICAL APPENDIX FOR CHAPTER 4

B.1 Background

We consider the following dynamic panel spatial model which is the combination of all the different piecewise framework which has been discussed in Chapter 4.

$$y_{it} = \gamma y_{it-1} + \tau \sum_{j=1}^{N} m_{ij} y_{jt} + \delta \sum_{j=1}^{N} m_{ij} y_{jt-1} + X_{it}\beta + u_{it} \tag{B.1}$$

$$u_{it} = \mu_i + \epsilon_{it} \tag{B.2}$$

$$\epsilon_{it} = \lambda \sum_{j=1}^{N} w_{ij} \epsilon jt + v_{it} \tag{B.3}$$

$$v_{it} = \rho v_{it-1} + e_{it}, \text{where } e_{it} \sim IIDN(0, \sigma_e^2) \tag{B.4}$$

for $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$. Here y_{it} is the observation for i^{th} individual/observation at t^{th} time, X_{it} denotes the observations on non-stochastic regressors and u_{it} is the regression disturbance. Spatial dependence is captured by the weigh matrices $M = (m_{ij})$ and $W = (w_{ij})$. In this framework, I have considered spatial lag dependence (τ), time dynamics (γ), space recursive (δ), spatial error dependence (λ), serial correlation in error (ρ) and individual effect (μ_i). I consider random effect model here, i.e., $\mu_i \sim IID(0, \sigma_\mu)$, like Sen and Bera (2011). W and M are row-standardized weight matrices whose diagonal elements are zero, such that $(I - \tau M)$ and $(I - \lambda W)$ are non-singular, where I is an identity matrix of dimension N. I assume that the model satisfies the regularity conditions given in Lee and Yu (2010).

In matrix form, the equations (B.1) - (B.4) can be rewritten as

$$y = \tau(I_T \otimes M)y + [(\gamma + \delta M) \otimes I_T]ly + X\beta + u \tag{B.5}$$

where y is f dimension $NT \times 1, X$ is $NT \times K, \beta$ is $k \times 1$ and u is $NT \times 1$. Here l is the lag operator, X is assumed to be of full column rank and its elements are bounded in absolute value. The disturbance term can be expressed as

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})v \tag{B.6}$$

where $B = (I_N - \lambda W), \iota_T$ is vector of ones of dimension T, I_T is an identity matrix of dimension $T \times T$ and $\otimes$ denotes Kronecker product. Under this setup, the variance-covariance matrix of u is given by

$$\Omega = \sigma_{\mu^2}[J_T \otimes I_N] + [V \otimes (B'B)^{-1}] \tag{B.7}$$

where J_T is a matrix of ones of dimension $T \times T$, and V is the familiar $T \times T$ variance -covariance matrix for AR (1) process in equation (B.4), i.e.,

$$V = E(v'v) = [\frac{1}{1-\rho^2}V_1] \otimes \sigma_e^2 I_N = V_\rho \otimes \sigma_e^2 I_N \tag{B.8}$$

with

$$V_1 = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \vdots & \vdots & \ddots & \vdots & \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 & \end{bmatrix}$$

and $V_\rho = \frac{1}{1-\rho^2}V_1$.

The loglikelihood function of the above model can be written as:

$$\begin{aligned} L = \frac{-NT}{2} ln2\pi - \frac{1}{2}\ln|\Omega| + T\ln|A| - \frac{1}{2}[(I_T \otimes A)y \\ - [(\gamma + \delta M) \otimes I_T]ly - X\beta]'\Omega^{-1}[(I_T \otimes A)y - [(\gamma + \delta M) \otimes I_T]ly - X\beta] \end{aligned} \tag{B.9}$$

where $A = (I_N - \tau M)$ and following Sen and Bera (2011), I can write

$$\frac{1}{2}\ln|\Omega| = -\frac{N}{2}\ln(1-\rho^2) + \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| + \frac{NT}{2}\ln\sigma_e^2 - (T-1)\ln|B|$$

where $d^2 = \alpha^2 + (T-1)$, $\alpha = \sqrt{\frac{1+\rho}{1-\rho}}$ and $\phi = \frac{\sigma_\mu^2}{\sigma_e^2}$. Thus substituting $\frac{1}{2}\ln|\Omega|$ in L , I obtain

$$\begin{aligned} L = \frac{-NT}{2}ln2\pi + \frac{N}{2}\ln(1-\rho^2) - \frac{1}{2}\ln|d^2(1-\rho)^2\phi I_N + (B'B)^{-1}| - \frac{NT}{2}\ln\sigma_e^2 + \\ (T-1)\ln|B| + T\ln|A| - \frac{1}{2}[(I_T \otimes A)y - [(\gamma+\delta M)\otimes I_T]ly - X\beta]' \\ \Omega^{-1}[(I_T \otimes A)y - [(\gamma+\delta M)\otimes I_T]ly - X\beta] \end{aligned} \tag{B.10}$$

B.2 Derivation of Score

$$\frac{\partial L}{\partial \beta} = X'\Omega^{-1}u \tag{B.11}$$

$$\frac{\partial L}{\partial \sigma_e^2} = -\frac{1}{2}trC^{-1}\frac{(d^2(1-\rho)^2\sigma_\mu^2 I_N}{\sigma_e^4} - \frac{NT}{2\sigma_e^2} - \frac{1}{2}u'(\Omega^{-1}(V_\rho \otimes [(B'B)^{-1}])\Omega^{-1})u \tag{B.12}$$

$$\frac{\partial L}{\partial \gamma} = (I_T \otimes Y_{NT-1})'\Omega^{-1}u \tag{B.13}$$

$$\frac{\partial L}{\partial \sigma_\mu^2} = -\frac{1}{2}trC^{-1}\frac{(d^2(1-\rho)^2 I_N}{\sigma_e^2} + \frac{1}{2}u'\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \tag{B.14}$$

$$\begin{aligned} \frac{\partial L}{\partial \rho} = -\frac{N\rho}{1-\rho^2} + \frac{1}{2}trC^{-1}(\rho + (T-1)(1-\rho)\phi I_N) \\ + \frac{1}{2}u'(\sigma_e^{-2}(\frac{1}{1-\rho^2})^2[2\rho V_1 + (1-\rho^2)F_\rho] \otimes (B'B)^{-1})u \end{aligned} \tag{B.15}$$

$$\frac{\partial L}{\partial \delta} = [(W \otimes I_T)Y_{NT-1}]\Omega^{-1}u \tag{B.16}$$

$$\frac{\partial L}{\partial \tau} = -Ttr(A^{-1}W) + \frac{1}{2}\Omega^{-1}(I_T \otimes W)y \tag{B.17}$$

$$\frac{\partial L}{\partial \lambda} = -(T-1)tr(B^{-1}W) + \frac{1}{2}trC^{-1}[(B'B)^{-1}[B'W+W'B](B'B)^{-1}] - \frac{1}{2}u'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{B.18}$$

where $C = (d^2(1-\rho)^2\phi I_N + (B'B)^{-1})$. The score functions evaluated under H_0^a, i.e., restricted MLE of θ_0 with $\tilde{\omega} = (\tilde{\beta}, \tilde{\sigma_e^2})$ are:

$$\frac{\partial L}{\partial \beta} = 0 \tag{B.19}$$

$$\frac{\partial L}{\partial \sigma_e^2} = 0 \tag{B.20}$$

$$\frac{\partial L}{\partial \gamma} = \frac{[I_T \otimes Y_{NT-1}]\tilde{u}'}{\sigma_e^2} \tag{B.21}$$

$$\frac{\partial L}{\partial \sigma_\mu^2} = \frac{NT}{2\tilde{\sigma_e^2}}[\frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1] \tag{B.22}$$

$$\frac{\partial L}{\partial \rho} = \frac{NT}{2}[\frac{\tilde{u}'(G \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{B.23}$$

$$\frac{\partial L}{\partial \delta} = \frac{\tilde{u}'[(I_T \otimes W)Y_{NT-1}]}{\tilde{\sigma_e^2}} \tag{B.24}$$

$$\frac{\partial L}{\partial \tau} = \frac{\tilde{u}'[(I_T \otimes W)Y_{NT}]}{\tilde{\sigma_e^2}} \tag{B.25}$$

$$\frac{\partial L}{\partial \lambda} = \frac{NT}{2}[\frac{\tilde{u}'(I_T \otimes (W+W'))\tilde{u}}{\tilde{u}'\tilde{u}}] \tag{B.26}$$

where $\tilde{u} = y - x\tilde{\beta}$ is the OLS residual vector, and $\tilde{\sigma_e^2} = \frac{\tilde{u}'\tilde{u}}{NT}$.

B.3 Derivation of Information Matrix

To derive the information matrix under joint null H_o^a, I need to derive the second order derivatives and take expectation. Differentiating equation (B.11) wrt $\beta, \sigma_e^2, \gamma, \sigma_\mu^2, \rho, \delta, \tau$ and λ, we get

$$\frac{\partial^2 L}{\partial \beta \partial \beta'} = -X'\Omega^{-1}X \tag{B.27}$$

$$\frac{\partial^2 L}{\partial\beta\partial\sigma_e^2} = -u'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}X \tag{B.28}$$

$$\frac{\partial^2 L}{\partial\beta\partial\gamma} = -(I_T \otimes Y_{NT-1})'\Omega^{-1}X \tag{B.29}$$

$$\frac{\partial^2 L}{\partial\beta\partial\sigma_\mu^2} = -X'\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \tag{B.30}$$

$$\frac{\partial^2 L}{\partial\beta\partial\rho} = -X'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{B.31}$$

$$\frac{\partial^2 L}{\partial\beta\partial\delta} = -[(W \otimes I_T)Y_{NT-1}]'\Omega^{-1}X \tag{B.32}$$

$$\frac{\partial^2 L}{\partial\beta\partial\tau} = -X'\Omega^{-1}(I_T \otimes W)Y \tag{B.33}$$

$$\frac{\partial^2 L}{\partial\beta\partial\lambda} = -X'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{B.34}$$

Differentiating equation (B.12) wrt $\gamma, \sigma_\mu^2, \rho, \delta, \tau$ and λ, we get

$$\frac{\partial^2 L}{\partial\sigma_e^2\partial\gamma} = -(I_T \otimes Y_{NT-1})\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{B.35}$$

$$\frac{\partial^2 L}{\partial\sigma_e^2\partial\sigma_\mu^2} = \frac{1}{2}tr[C^{-1}\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}C^{-1}(B'B)^{-1}] - u\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{B.36}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial\sigma_e^2\partial\rho} &= -tr[C^{-1}[(\rho + (1-\rho)(T-1))\phi I_N]C^{-1}(B'B)^{-1}] \\ &\quad - u\Omega^{-1}[[2\frac{\rho}{1-\rho^2}^2 V_1 + \frac{1}{1-\rho^2}F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \\ &\quad + \frac{1}{2}u\Omega^{-1}[[\frac{2\rho}{(1-\rho^2)^2}V_1 + \frac{1}{1-\rho^2}F_\rho] \otimes (B'B)^{-1}]\end{aligned} \tag{B.37}$$

$$\frac{\partial^2 L}{\partial\sigma_e^2\partial\delta} = -((W \otimes I_T)Y_{NT-1})'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{B.38}$$

$$\frac{\partial^2 L}{\partial\sigma_e^2\partial\tau} = -((W \otimes I_T)Y_{NT})'\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u \tag{B.39}$$

$$\frac{\partial^2 L}{\partial \sigma_e^2 \partial \lambda} = -\frac{1}{2}tr[C^{-1}(B'B)^{-1}(B'W + W'B)(B'B)^{-1}C^{-1}(B'B)^{-1} + C^{-1}(B'B)^{-1}$$
$$(B'W + W'B)(B'B)^{-1}] - \frac{T-1}{2}tr[(B'B)^{-1}(B'W + W'B)(B'B)^{-1}] - u\Omega^{-1}[V_\rho$$
$$\otimes (B'B)^{-1}[B'W + W'B](B'B)^{-1}]\Omega^{-1}(V_\rho \otimes (B'B)^{-1})\Omega^{-1}u + \frac{1}{2}u\Omega^{-1}$$
$$(V_\rho \otimes (B'B)^{-1}(B'W + W'B)(B'B)^{-1})\Omega^{-1}u \quad \text{(B.40)}$$

Differentiating equation (B.13) wrt $\gamma, \sigma_\mu^2, \rho, \delta, \tau$ and λ, we get

$$\frac{\partial^2 L}{\partial \gamma \partial \gamma} = -(I_T \otimes Y_{NT-1})'\Omega^{-1}(I_T \otimes Y_{NT-1}) \quad \text{(B.41)}$$

$$\frac{\partial^2 L}{\partial \gamma \partial \sigma_\mu^2} = -(I_T \otimes Y_{NT-1})'\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \quad \text{(B.42)}$$

$$\frac{\partial^2 L}{\partial \gamma \partial \rho} = -(I_T \otimes Y_{NT-1})'\Omega^{-1}[[\frac{2\rho}{(1-\rho^2)^2}V_1 + \frac{1}{(1-\rho^2)}F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}u \quad \text{(B.43)}$$

$$\frac{\partial^2 L}{\partial \gamma \partial \delta} = -(I_T \otimes Y_{NT-1})'\Omega^{-1}[(W \otimes I_T)Y_{NT-1}] \quad \text{(B.44)}$$

$$\frac{\partial^2 L}{\partial \gamma \partial \tau} = -(I_T \otimes Y_{NT-1})'\Omega^{-1}[(W \otimes I_T)Y_{NT}] \quad \text{(B.45)}$$

$$\frac{\partial^2 L}{\partial \gamma \partial \lambda} = -(I_T \otimes Y_{NT-1})'\Omega^{-1}[V_\rho \otimes (B'B)^{-1}(B'W + W'B)(B'B)^{-1}]\Omega^{-1}u \quad \text{(B.46)}$$

Differentiating equation (B.14) wrt $\sigma_\mu^2, \rho, \delta, \tau$ and λ, we get

$$\frac{\partial^2 L}{\partial \sigma_\mu^2 \partial \sigma_\mu^2} = \frac{1}{2}tr[C^{-1}\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}C^{-1}\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}] - u\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \quad \text{(B.47)}$$

$$\frac{\partial^2 L}{\partial \sigma_\mu^2 \partial \rho} = \frac{1}{2}tr[\frac{d^2(1-\rho)^2 I_N}{\sigma_e^2}[C^{-1}(\rho + (T-1)(1-\rho)\phi I_N C^{-1})] + u'\Omega^{-1}[\sigma_e^2(\frac{1}{1-\rho^2})^2[2\rho V_1$$
$$+ (1-\rho^2)F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \quad \text{(B.48)}$$

$$\frac{\partial^2 L}{\partial \sigma_\mu^2 \partial \delta} = -[(W \otimes I_T)Y_{NT-1}]'\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \quad \text{(B.49)}$$

$$\frac{\partial^2 L}{\partial \sigma_\mu^2 \partial \tau} = -[(W \otimes I_T)Y_{NT}]'\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \tag{B.50}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \sigma_\mu^2 \partial \lambda} &= -\frac{1}{2}tr[\frac{d^2(1-\rho)^2}{\sigma_e^2}I_N C^{-1}\frac{d^2(1-\rho)^2}{\sigma_e^2}I_N C^{-1}] \\ &+ u'\Omega^{-1}[\sigma_e^2(V_\rho \otimes [(B'B)^{-1}[B'W + W'B](B'B)^{-1}]\Omega^{-1}(J_T \otimes I_N)\Omega^{-1}u \end{aligned}\tag{B.51}$$

Differentiating equation (B.15) wrt ρ, δ, τ and λ, we get

$$\begin{aligned}\frac{\partial^2 L}{\partial \rho \partial \rho} &= \frac{-N + N\rho^2}{(1-\rho^2)^2} + \frac{1}{2}tr((2\rho + (T-1)(1-\rho))\phi I_N) \\ &[C^{-1}(\rho + (T-1)(1-\rho))\phi I_N C^{-1}] + u\Omega^{-1}[\sigma_e^2(\frac{1}{1-\rho^2})^2[2\rho V_1 + (1-\rho^2)F_\rho] \\ &\otimes (B'B)^{-1}]\Omega^{-1}[\sigma_e^2(\frac{1}{1-\rho^2})^2[2\rho V_1 + (1-\rho^2)F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}u + u \\ &\Omega^{-1}[[\sigma_e^2(\frac{1}{1-\rho^2})^2[2V_1 - 2\rho F_\rho] + 4(1-\rho^2)\rho(2\rho V_1 \\ &+ (1-\rho^2)F_\rho) \otimes (B'B)^{-1}]\Omega^{-1}u \end{aligned}\tag{B.52}$$

$$\frac{\partial^2 L}{\partial \rho \partial \delta} = -[(W \otimes I_T)Y_{NT-1}]\Omega^{-1}[[\frac{2\rho}{(1-\rho^2)^2}V_1 + \frac{1}{1-\rho^2}F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}u \tag{B.53}$$

$$\frac{\partial^2 L}{\partial \rho \partial \tau} = -[(W \otimes I_T)Y_{NT}]\Omega^{-1}[[\frac{2\rho}{(1-\rho^2)^2}V_1 + \frac{1}{1-\rho^2}F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}u \tag{B.54}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \rho \partial \lambda} &= \frac{1}{2}tr[[(B'B)^{-1}[B'W + W'B](B'B)^{-1}C6{-}1(\rho + (T-1)(1-\rho))\phi I_N C^{-1}] \\ &+ u\Omega^{-1}[\sigma_e^2(\frac{1}{1-\rho^2})^2[2V_1 - 2\rho F_\rho] \otimes (B'B)^{-1}]\Omega^{-1}[\sigma_e^2(V_\rho \otimes [(B'B)^{-1} \\ &[B'W + W'B](B'B)^{-1}]\Omega^{-1} - \frac{1}{2}u\Omega^{-1}[\sigma_e^4(\frac{1}{1-\rho^2})^2[2V_1 - 2\rho F_\rho]\otimes \\ &[(B'B)^{-1}[B'W + W'B](B'B)^{-1}](B'B)^{-1}]\Omega^{-1}u \end{aligned}\tag{B.55}$$

Differentiating equation (B.16) wrt δ, τ and λ, we get

$$\frac{\partial^2 L}{\partial \delta \partial \delta} = -[(W \otimes I_T)Y_{NT-1}]'\Omega^{-1}[(W \otimes I_T)Y_{NT-1}] \tag{B.56}$$

$$\frac{\partial^2 L}{\partial\delta\partial\tau} = -[(W\otimes I_T)Y_{NT-1}]'\Omega^{-1}[(W\otimes I_T)Y_{NT}] \tag{B.57}$$

$$\frac{\partial^2 L}{\partial\delta\partial\lambda} = -[(W\otimes I_T)Y_{NT-1}]'\Omega^{-1}(V_\rho(B'B)^{-1}(W'B+B'W)(B'B)^{-1})\Omega^{-1}u \tag{B.58}$$

Differentiating equation (B.17) wrt τ and λ, we get

$$\frac{\partial^2 L}{\partial\tau\partial\tau} = -Ttr((A^{-1}W)^2) - Y_{NT}(I_T\otimes W)\Omega^{-1}(I_T\otimes W)Y_{NT} \tag{B.59}$$

$$\frac{\partial^2 L}{\partial\tau\partial\lambda} = -u\Omega^{-1}(V_\rho\otimes(B'B)^{-1})\Omega^{-1}(I_T\otimes W)u \tag{B.60}$$

For $\frac{\partial^2 L}{\partial\lambda\partial\lambda}$ please see Appendix A.

Under the joint null $H_o^a : \gamma = \sigma_\mu^2 = \rho = \delta = \tau = \lambda = 0$, the non-zero second-order derivatives are :

$\frac{\partial^2 L}{\partial\beta\partial\beta} = -\frac{X'X}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\beta\partial\gamma} = -\frac{(I_T\otimes Y_{NT-1})'X}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\beta\partial\delta} = -\frac{[(W\otimes I_T)Y_{NT-1}]'X}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\beta\partial\tau} = -\frac{X'(W\otimes I_T)X\hat{\beta}}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\sigma_e^2\partial\sigma_e^2} = -\frac{NT}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\sigma_e^2\partial\gamma} = -\frac{(I_T\otimes Y_{NT-1})(I_T\otimes I_N)u}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\sigma_e^2\partial\sigma_\mu^2} = -\frac{NT}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\gamma\partial\gamma} = -\frac{(I_T\otimes Y_{NT-1})'(I_T\otimes Y_{NT-1})}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\gamma\partial\sigma_\mu^2} = -\frac{(I_T\otimes Y_{NT-1})'(J_T\otimes I_N)u}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\gamma\partial\rho} = -\frac{(I_T\otimes Y_{NT-1})'(I_T\otimes I_N)u}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\sigma_\mu^2} = -\frac{NT^2}{2\hat{\sigma_e^4}}$
$\frac{\partial^2 L}{\partial\sigma_\mu^2\partial\rho} = -\frac{N(T-1)}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\rho\partial\rho} = -N(T-1)$
$\frac{\partial^2 L}{\partial\delta\partial\delta} = -\frac{((W\otimes I_T)Y_{NT-1})'((W\otimes I_T)Y_{NT-1})}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\delta\partial\tau} = -\frac{((W\otimes I_T)Y_{NT-1})'((W\otimes I_T)Y_{NT-1})}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\delta\partial\lambda} = -\frac{((W\otimes I_T)Y_{NT-1})'((I_T\otimes(W+W'))u}{\hat{\sigma_e^2}}$
$\frac{\partial^2 L}{\partial\tau\partial\tau} = -(Ttr(W^2+WW') + \frac{(\hat{\beta}'X'(I_T\otimes W')(I_T\otimes W)X\hat{\beta})}{\hat{\sigma_e^2}})$

$\frac{\partial^2 L}{\partial\tau\partial\lambda} = \frac{\partial^2 L}{\partial\lambda\partial\lambda} = -Ttr(W^2 + WW')$.

All the other second derivatives becomes zero under joint null. Thus the information matrix J, under H_o^a is

$$J(\theta_o) = \begin{bmatrix} J_\beta & 0 & J_{\beta\gamma} & 0 & 0 & J_{\beta\delta} & J_{\beta\tau} & 0 \\ 0 & J_{\sigma_e^2} & J_{\sigma_e^2\gamma} & J_{\sigma_e^2\sigma_\mu^2} & 0 & 0 & 0 & 0 \\ J_{\gamma\beta} & J_{\gamma\sigma_e^2} & J_\gamma & J_{\gamma\sigma_\mu^2} & J_{\gamma\rho} & 0 & 0 & 0 \\ 0 & J_{\sigma_\mu^2\sigma_e^2} & J_{\sigma_\mu^2\gamma} & J_{\sigma_\mu^2} & J_{\sigma_\mu^2\rho} & 0 & 0 & 0 \\ 0 & 0 & J_{\rho\gamma} & J_{\rho\sigma_\mu^2} & J_\rho & 0 & 0 & 0 \\ J_{\delta\beta} & 0 & 0 & 0 & 0 & J_\delta & J_{\delta\tau} & J_{\delta\lambda} \\ J_{\tau\beta} & 0 & 0 & 0 & 0 & J_{\tau\delta} & J_\tau & J_{\tau\lambda} \\ 0 & 0 & 0 & 0 & 0 & J_{\lambda\delta} & J_{\lambda\tau} & J_\lambda \end{bmatrix} \quad \text{(B.61)}$$

where $J = E(-\frac{1}{NT}\frac{\partial^2 L}{\partial\tilde{\theta}\partial\tilde{\theta}'})$ evaluated at θ_0.

B.4 Derivation of test statistics

Recall from Chapter 4, Section 4.3:

$$RS_\psi^* = \frac{1}{N}[d_\psi(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})d_\phi(\tilde{\theta})][J_{\psi.\omega}(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})J_{\phi\psi.\omega}(\tilde{\theta})]^{-1}$$
$$[d_\psi(\tilde{\theta}) - J_{\psi\phi.\omega}(\tilde{\theta})J_{\phi.\omega}^{-1}(\tilde{\theta})d_\phi(\tilde{\theta})]' \quad \text{(B.62)}$$

where $\omega = (\beta', \sigma_e^2)'$, ψ and ϕ are different combinations of the parameters $(\gamma, \sigma_\mu^2, \rho, \delta, \tau, \lambda)$.
I) $H_o^b : \gamma = 0$ in presence of $\phi = (\sigma_\mu^2, \rho, \delta, \tau, \lambda)$.
Here we are testing the significance of time-dynamics γ, in presence of random effects, serial correlation, and spatial dependence.
$d_\psi = d_\gamma$
$d_\phi = (d_{\sigma_\mu^2}, d_\rho, d_\delta, d_\tau, d_\lambda)$
$J_{\psi\phi.\omega} = J_{\psi\phi} - J_{\psi\omega}J_\omega^{-1}J_{\phi\omega} = (J_{\gamma\sigma_\mu^2.\sigma_e^2}, J_{\gamma\rho}, 0, 0, 0)$

$J_{\phi.\omega} = J_\phi - J_{\phi\omega} J_\omega^{-1} J_{\omega\phi} =$

$$\begin{bmatrix} J_{\sigma_\mu^2.\sigma_e^2} & J_{\sigma_\mu^2\rho} & 0 & 0 & 0 \\ J_{\rho\sigma_\mu^2} & J_\rho & 0 & 0 & 0 \\ 0 & 0 & J_{\delta.\beta} & J_{\delta\tau.\beta} & J_{\delta\lambda} \\ 0 & 0 & J_{\tau\delta.\beta} & J_{\tau.\beta} & J_{\tau\lambda} \\ 0 & 0 & J_{\lambda\delta} & J_{\lambda\tau} & J_\lambda \end{bmatrix}$$

Therefore, adjusted proposed test statistic for time-dynamics γ is:

$$RS^*_\gamma = [d_\gamma - J_{\gamma\sigma_\mu^2.\sigma_e^2} J^{-1}_{\sigma_\mu^2.\sigma_e^2} d_{\sigma_\mu^2} - J_{\gamma\rho} J_\rho^{-1} d_\rho][J_{\gamma.\omega} - J_{\gamma\sigma_\mu^2.\sigma_e^2} J^{-1}_{\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2\gamma.\sigma_e^2} - J_{\gamma\rho} J_\rho^{-1} J_{\rho\gamma}]^{-1}$$
$$[d_\gamma - J_{\gamma\sigma_\mu^2.\sigma_e^2} J^{-1}_{\sigma_\mu^2.\sigma_e^2} d_{\sigma_\mu^2} - J_{\gamma\rho} J_\rho^{-1} d_\rho]' \to \chi_1^2(0). \quad \text{(B.63)}$$

II) $H_o^c : \sigma_\mu^2 = 0$ in presence of $\gamma, \rho, \delta, \tau, \lambda$.

Here $\phi = (\gamma, \rho, \delta, \tau, \lambda)$

$d_\psi = d_{\sigma_{\mu^2}}$

$d_\phi = (d_\gamma, d_\rho, d_\delta, d_\tau, d_\lambda)$

$J_{\psi\phi.\omega} = (J_{\sigma_\mu^2\gamma.\sigma_e^2}, J_{\sigma_\mu^2\rho}, 0, 0, 0)$

$$J_{\phi.\omega} = \begin{bmatrix} J_{\gamma.\omega} & J_{\gamma\rho} & 0 & 0 & 0 \\ J_{\rho\gamma} & J_\rho & 0 & 0 & 0 \\ 0 & 0 & J_{\delta.\beta} & J_{\delta\tau.\beta} & J_{\delta\lambda} \\ 0 & 0 & J_{\tau\delta.\beta} & J_{\tau.\beta} & J_{\tau\lambda} \\ 0 & 0 & J_{\lambda\delta} & J_{\lambda\tau} & J_\lambda \end{bmatrix}$$

The adjusted RS test statistics is:

$$RS^*_{\sigma_\mu^2} = [d_{\sigma_\mu^2} - J_{\sigma_\mu^2\gamma.\sigma_e^2} J^{-1}_{\gamma.\omega} d_\gamma - J_{\sigma_\mu^2\rho} J_\rho^{-1} d_\rho][J_{\sigma_\mu^2.\sigma_e^2} - J_{\sigma_\mu^2\gamma.\sigma_e^2} J^{-1}_{\gamma.\omega} J_{\gamma\sigma_\mu^2.\sigma_e^2} - J_{\sigma_\mu^2\rho} J_\rho^{-1} J_{\rho\sigma_\mu^2}]^{-1}$$
$$[d_{\sigma_\mu^2} - J_{\sigma_\mu^2\gamma.\sigma_e^2} J^{-1}_{\gamma.\omega} d_\gamma - J_{\sigma_\mu^2\rho} J_\rho^{-1} d_\rho]' \to \chi_1^2(0), \quad \text{(B.64)}$$

III) $H_o^d : \rho = 0$ in presence of $\gamma, \sigma_\mu^2, \delta, \tau, \lambda$.

Here $\phi = (\gamma, \sigma_\mu^2, \delta, \tau, \lambda)$.

$d_\psi = d_\rho$

$d_\phi = (d_\gamma, d_{\sigma_{\mu^2}}, d_\delta, d_\tau, d_\lambda)$

$J_{\psi\phi.\omega} = (J_{\rho\gamma}, J_{\rho\sigma_\mu^2}, 0, 0, 0)$.

$$J_{\phi.\omega} = \begin{bmatrix} J_{\gamma.\omega} & J_{\gamma\sigma_\mu^2.\sigma_e^2} & 0 & 0 & 0 \\ J_{\sigma_\mu^2\gamma.\sigma_e^2} & J_{\sigma_\mu^2.\sigma_e^2} & 0 & 0 & 0 \\ 0 & 0 & J_{\delta.\beta} & J_{\delta\tau.\beta} & J_{\delta\lambda} \\ 0 & 0 & J_{\tau\delta.\beta} & J_{\tau.\beta} & J_{\tau\lambda} \\ 0 & 0 & J_{\lambda\delta} & J_{\lambda\tau} & J_\lambda \end{bmatrix}$$

The adjusted test statistic is:

$$RS_\rho^* = [d_\rho - J_{\rho\gamma.\sigma_e^2} J_{\gamma.\omega}^{-1} d_\gamma - J_{\rho\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2.\sigma_e^2}^{-1} d_{\sigma_\mu^2}][J_\rho - J_{\rho\gamma.\sigma_e^2} J_{\gamma.\omega}^{-1} J_{\gamma\rho.\sigma_e^2} - J_{\rho\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2.\sigma_e^2}^{-1} J_{\sigma_\mu^2\rho.\sigma_e^2}]^{-1}$$
$$[d_\rho - J_{\rho\gamma.\sigma_e^2} J_{\gamma.\omega}^{-1} d_\gamma - J_{\rho\sigma_\mu^2.\sigma_e^2} J_{\sigma_\mu^2.\sigma_e^2}^{-1} d_{\sigma_\mu^2}]' \to \chi_1^2 \quad \text{(B.65)}$$

IV) $H_o^e : \delta = 0$ in presence of $\gamma, \sigma_\mu^2, \rho, \tau, \lambda$.

Here $\phi = (\gamma, \sigma_\mu^2, \rho, \tau, \lambda)$.

$d_\psi = d_\delta$

$d_\phi = (d_\gamma, d_{\sigma_{\mu^2}}, d_\rho, d_\tau, d_\lambda)$

$J_{\psi\phi.\omega} = (0, 0, 0, J_{\delta\lambda}, J_{\delta\tau.\beta})$.

$$J_{\phi.\omega} = \begin{bmatrix} J_{\gamma.\omega} & J_{\gamma\sigma_\mu^2.\sigma_e^2} & J_{\gamma\rho} & 0 & 0 \\ J_{\sigma_\mu^2\gamma.\sigma_e^2} & J_{\sigma_\mu^2.\sigma_e^2} & J_{\sigma_\mu^2\rho} & 0 & 0 \\ J_{\rho\gamma} & J_{\rho\sigma_\mu^2} & J_\rho & 0 & 0 \\ 0 & 0 & 0 & J_{\tau.\beta} & J_{\tau\lambda} \\ 0 & 0 & 0 & J_{\lambda\tau} & J_\lambda \end{bmatrix}$$

The test statistic for space recursive parameter δ is directly affected by the other spatial parameters λ and τ, and not by other parameters. The separation between the spatial parameters and all the other parameters is also distinct here, as far as the test statistic is concerned. The adjusted RS test statistic is:

$$RS^*_\delta = [d_\delta - J_{(\delta\lambda).(\tau.\beta)} J^{-1}_{\lambda.(\tau.\beta)} d_\lambda - J_{(\delta\tau.\beta).\lambda} J^{-1}_{(\tau.\beta).\lambda} d_\tau]$$
$$[J_\delta - J_{(\delta\lambda).(\tau.\beta)} J^{-1}_{\lambda.(\tau.\beta)} J_{(\lambda\delta).(\tau.\beta)} - J_{(\delta\tau.\beta).\lambda} J^{-1}_{(\tau.\beta).\lambda} J_{(\tau\delta.\beta).\lambda}]^{-1}$$
$$[d_\delta - J_{(\delta\lambda).(\tau.\beta)} J^{-1}_{\lambda.(\tau.\beta)} d_\lambda - J_{(\delta\tau.\beta).\lambda} J^{-1}_{(\tau.\beta).\lambda} d_\tau]' \sim \chi^2_1(0) \quad \text{(B.66)}$$

V) $H^f_o : \tau = 0$ in presence of $\gamma, \sigma^2_\mu, \rho, \delta, \lambda$.

Here $\phi = (\gamma, \sigma^2_\mu, \rho, \delta, \lambda)$

$d_\psi = d_\tau$

$d_\phi = (d_\gamma, d_{\sigma_{\mu^2}}, d_\rho, d_\delta, d_\lambda)$

$J_{\psi\phi.\omega} = (0, 0, 0, J_{\tau\delta.\beta}, J_{\tau\lambda})$.

$$J_{\phi.\omega} \begin{bmatrix} J_{\gamma.\omega} & J_{\gamma\sigma^2_\mu.\sigma^2_e} & J_{\gamma\rho} & 0 & 0 \\ J_{\sigma^2_\mu\gamma.\sigma^2_e} & J_{\sigma^2_\mu.\sigma^2_e} & J_{\sigma^2_\mu\rho} & 0 & 0 \\ J_{\rho\gamma} & J_{\rho\sigma^2_\mu} & J_\rho & 0 & 0 \\ 0 & 0 & 0 & J_{\delta.\tau} & J_{\delta\lambda} \\ 0 & 0 & 0 & J_{\lambda\delta} & J_\lambda \end{bmatrix}$$

The adjusted test statistic is:

$$RS^*_\tau = [d_\tau - J_{(\tau\delta.\beta).\lambda} J^{-1}_{(\delta.\beta).\lambda} d_\delta - J_{\tau\lambda.(\delta.\beta)} J^{-1}_{\lambda.(\delta.\beta)} d_\lambda]$$
$$[J_{\tau.\beta} - J_{(\tau\delta.\beta).\lambda} J^{-1}_{(\delta.\beta).\lambda} J_{(\delta\tau.\beta).\lambda} - J_{\tau\lambda.(\delta.\beta)} J^{-1}_{\lambda.(\delta.\beta)} J_{\lambda\tau.(\delta.\beta)}]^{-1}$$
$$[d_\tau - J_{(\tau\delta.\beta).\lambda} J^{-1}_{(\delta.\beta).\lambda} d_\delta - J_{\tau\lambda.(\delta.\beta)} J^{-1}_{\lambda.(\delta.\beta)} d_\lambda]' \sim \chi^2_1(0) \quad \text{(B.67)}$$

Lastly, VI) $H^g_o : \lambda = 0$ in presence of $\gamma, \sigma^2_\mu, \rho, \delta, \tau$.

Here, $\phi = (\gamma, \sigma^2_\mu, \rho, \delta, \tau)$

$d_\psi = d_\lambda$

$d_\phi = (d_\gamma, d_{\sigma_{\mu^2}}, d_\rho, d_\delta, d_\tau)$

$J_{\psi\phi.\omega} = (0, 0, 0, J_{\lambda\delta}, J_{\lambda\tau})$.

$$J_{\phi.\omega} \begin{bmatrix} J_{\gamma.\omega} & J_{\gamma\sigma_\mu^2.\sigma_e^2} & J_{\gamma\rho} & 0 & 0 \\ J_{\sigma_\mu^2\gamma.\sigma_e^2} & J_{\sigma_\mu^2.\sigma_e^2} & J_{\sigma_\mu^2\rho} & 0 & 0 \\ J_{\rho\gamma} & J_{\rho\sigma_\mu^2} & J_\rho & 0 & 0 \\ 0 & 0 & 0 & J_{\delta.\beta} & J_{\delta\tau.\beta} \\ 0 & 0 & 0 & J_{\tau\delta.\beta} & J_{\tau.\beta} \end{bmatrix}$$

The adjusted test statistic is:

$$\begin{aligned} RS_\lambda^* = [d_\lambda - J_{(\lambda\delta).(\tau.\beta)} J^{-1}_{(\delta.\beta).(\tau.\beta)} d_\delta - J_{\lambda\tau.(\delta.\beta)} J^{-1}_{(\tau.\beta).(\delta.\beta)} d_\tau][J_\lambda - J_{(\lambda\delta).(\tau.\beta)} J^{-1}_{(\delta.\beta)(\tau.\beta)} J_{(\delta\lambda).(\tau.\beta)} \\ - J_{\lambda\tau.\delta} J^{-1}_{(\tau.\beta).(\delta.\beta)} J_{\tau\lambda.(\delta.\beta)}]^{-1} [d_\lambda - J_{(\lambda\delta).(\tau.\beta)} J^{-1}_{(\delta.\beta).(\tau.\beta)} d_\delta - J_{\lambda\tau.(\delta.\beta)} J^{-1}_{(\tau.\beta).(\delta.\beta)} d_\tau]' \sim \chi_1^2(0) \quad \text{(B.68)} \end{aligned}$$

B.5 Country Lists and Groups

Here are the list of countries in each group divided based on their initial income in 1961:

Panel A (Poorest): The average of real per-capita income has grown by 2.7% over 35 years. Bangladesh, Benin, Botswana, Burkina Faso, Central African Republic, Chad, Congo Dem. Rep., Ethiopia, Ghana, India, Indonesia, Kenya, Madagascar, Malawi, Mali, Mozambique, Nepal, Niger, Rwanda, Sierra Leone, Sri Lanka, Tanzania, Uganda, Zimbabwe.

Panel B (Middle): The growth rate of the real per capita income is 5.19% from 1961 - 1995. Angola, Bolivia, Cameroon, Republic of Congo, Cote d'Ivoire, Dominican Republic, Ecuador, Egypt, Honduras, Malaysia, Mauritania, Mauritius, Morocco, Nigeria, Pakistan, Papua New Guinea, Paraguay, Philippines, Senegal, Syria, Thailand, Tunisia, Zambia.

Panel C (Rich): The average of real per capita income has grown by 4.83 % over 35 years. Argentina, Brazil, Chile, Colombia, Costa Rica, El Salvador, Finland, Guatemala, Hong Kong, Ireland, Jamaica, Japan, Jordan, Korea, Mexico, Nicaragua, Panama, Peru, Portugal, Singapore, South Africa, Spain, Trinidad and Tobago, Turkey, Uruguay.

Panel D (Richest): The average growth rate of per capita income of this group is 4.37 %. Australia, Austria, Belgium, Canada, Denmark, France, Greece, Israel, Italy, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, United States, Venezuela.

Ranking of Countries according to income in 1995:

Poorest: Zimbabwe, Congo Dem. Rep., Burundi, Ethiopia, Central African Republic, Malawi, Mozambique, Madagascar, Niger, Togo, Rwanda, Burkina Faso, Tanzania, Sierra Leone, Ghana, Uganda, Nepal, Kenya, Bangladesh, Benin, Mali, Mauritania, Cote d'Ivoire, Chad, Senegal.

Middle: Zambia, Cameroon, Nigeria, Papua New Guinea, Republic of Congo, Nicaragua, Philippines, Pakistan, India, Angola, Indonesia, Bolivia, Paraguay, Sri Lanka, Morocco, Honduras, Syria, Thailand, Peru, Egypt, Ecuador, Jordan, Tunisia, Guatemala, El Salvador.

Rich: Brazil, Colombia, Panama, Dominican Republic, Uruguay, Mauritius, Botswana, South Africa, Venezuela, Jamaica, Argentina, Costa Rica, Malaysia, Turkey, Chile, Mexico, Portugal, Trinidad and Tobago, Korea, New Zealand, Spain, Israel, Greece, Japan, Finland.

Richest: United States, Italy, Belgium, Norway, Australia, Canada, France, Hong Kong, Netherlands, Denmark, Switzerland, Israel, Sweden, United Kingdom, Greece, Venezuela, Singapore, Ireland, Norway.

As can be seen from the lists and also evident from Figure 1 and its subsequent discussions, there has been some transitional changes among the groups.

B.6 More Monte Carlo Results

In addition to Table 4.3 - 4.4, in Section 4.7, I provide Tables B.1- B.4, in addition to support the good finite sample properties of the proposed tests. The tables are listed in the next page onwards.

Table B.1: Estimated Rejection Probabilities with $\delta = \tau = \lambda = 0$. Sample size: $N = 25, T = 10$

ρ	η	γ	RS_{γ}^{*}	RS_{γ}	$RS_{\sigma_{\mu}^{2}}^{*}$	$RS_{\sigma_{\mu}^{2}}$	RS_{ρ}^{*}	RS_{ρ}
0.0	0.2	0.0	0.066	0.955	0.043	0.099	0.072	0.166
0.1	0.2	0.0	0.030	0.997	0.071	0.111	0.112	0.231
0.2	0.2	0.0	0.075	1.000	0.053	0.205	0.192	0.312
0.3	0.2	0.0	0.042	1.000	0.135	0.214	0.274	0.401
0.4	0.2	0.0	0.054	1.000	0.313	0.358	0.440	0.521
0.5	0.2	0.0	0.034	1.000	0.203	0.189	0.712	0.682
0.0	0.0	0.2	0.107	1.000	0.037	0.107	0.035	0.105
0.1	0.0	0.2	0.305	1.000	0.045	0.123	0.127	0.108
0.2	0.0	0.2	0.293	1.000	0.053	0.253	0.120	0.118
0.3	0.0	0.2	0.331	1.000	0.091	0.147	0.209	0.499
0.4	0.0	0.2	0.401	1.000	0.061	0.311	0.219	0.881
0.5	0.0	0.2	0.554	1.000	0.051	0.556	0.293	0.994
0.0	0.0	0.4	0.611	1.000	0.041	0.031	0.051	0.182
0.1	0.0	0.4	0.602	1.000	0.042	0.082	0.104	0.336
0.2	0.0	0.4	0.742	1.000	0.061	0.141	0.202	0.692
0.3	0.0	0.4	0.739	1.000	0.038	0.231	0.284	0.944
0.4	0.0	0.4	0.779	1.000	0.049	0.419	0.335	0.995
0.5	0.0	0.4	0.891	0.999	0.001	0.627	0.207	1.000
0.0	0.2	0.2	0.105	1.000	0.184	0.207	0.035	0.191
0.1	0.2	0.2	0.204	1.000	0.304	0.313	0.125	0.302
0.2	0.2	0.2	0.305	1.000	0.305	0.545	0.112	0.351
0.3	0.2	0.2	0.312	1.000	0.502	0.790	0.201	0.565
0.4	0.2	0.2	0.441	1.000	0.589	0.811	0.304	0.628
0.5	0.2	0.2	0.618	1.000	0.612	0.812	0.419	0.920
0.2	0.0	0.0	0.082	1.000	0.056	0.159	0.097	0.118
0.2	0.1	0.0	0.077	1.000	0.062	0.312	0.109	0.212
0.2	0.2	0.0	0.093	0.998	0.164	0.302	0.118	0.318
0.2	0.3	0.0	0.092	0.999	0.161	0.316	0.199	0.401
0.2	0.4	0.0	0.075	0.998	0.251	0.515	0.100	0.399
0.2	0.5	0.0	0.094	0.998	0.298	0.532	0.109	0.333
0.0	0.0	0.2	0.107	1.000	0.060	0.117	0.032	0.092
0.0	0.1	0.2	0.118	1.000	0.097	0.105	0.036	0.108
0.0	0.2	0.2	0.204	1.000	0.104	0.209	0.024	0.112
0.0	0.3	0.2	0.204	1.000	0.203	0.306	0.022	0.308
0.0	0.4	0.2	0.307	1.000	0.335	0.413	0.036	0.410
0.0	0.5	0.2	0.406	1.000	0.416	0.508	0.028	0.399

Table B.2: Estimated Rejection Probabilities with $\delta = \tau = \lambda = 0$. Sample size: $N = 25, T = 10$

ρ	η	γ	RS_{γ}^{*}	RS_{γ}	$RS_{\sigma_\mu^2}^{*}$	$RS_{\sigma_\mu^2}$	RS_{ρ}^{*}	RS_{ρ}
0.2	0	0	0.036	0.845	0.041	0.103	0.104	0.301
0.2	0	0.1	0.330	0.990	0.031	0.115	0.147	0.306
0.2	0	0.2	0.275	1.000	0.040	0.151	0.114	0.213
0.2	0	0.3	0.201	1.000	0.051	0.184	0.102	0.394
0.2	0	0.4	0.311	1.000	0.037	0.123	0.105	0.682
0.2	0	0.5	0.314	1.000	0.045	0.152	0.143	0.846
0	0.2	0	0.052	0.970	0.162	0.206	0.042	0.112
0	0.2	0.1	0.141	1.000	0.133	0.301	0.073	0.212
0	0.2	0.2	0.206	1.000	0.205	0.335	0.026	0.102
0	0.2	0.3	0.331	1.000	0.201	0.421	0.051	0.211
0	0.2	0.4	0.312	1.000	0.333	0.533	0.059	0.181
0	0.2	0.5	0.415	1.000	0.398	0.546	0.071	0.253
0	0.4	0	0.041	0.990	0.152	0.550	0.058	0.123
0	0.4	0.1	0.142	1.000	0.225	0.601	0.043	0.233
0	0.4	0.2	0.106	1.000	0.105	0.607	0.027	0.341
0	0.4	0.3	0.204	1.000	0.204	0.715	0.043	0.311
0	0.4	0.4	0.301	1.000	0.301	0.681	0.043	0.368
0	0.4	0.5	0.411	1.000	0.312	0.747	0.039	0.221
0.2	0.2	0	0.070	1.000	0.049	0.302	0.100	0.230
0.2	0.2	0.1	0.212	1.000	0.117	0.315	0.139	0.101
0.2	0.2	0.2	0.307	1.000	0.107	0.449	0.216	0.141
0.2	0.2	0.3	0.301	1.000	0.201	0.568	0.206	0.256
0.2	0.2	0.4	0.412	1.000	0.312	0.621	0.301	0.561
0.2	0.2	0.5	0.512	1.000	0.376	0.632	0.303	0.770
0	0	0.4	0.328	1.000	0.041	0.110	0.057	0.033
0	0.1	0.4	0.399	1.000	0.197	0.215	0.034	0.178
0	0.2	0.4	0.458	1.000	0.104	0.311	0.042	0.108
0	0.3	0.4	0.502	1.000	0.233	0.311	0.047	0.271
0	0.4	0.4	0.555	1.000	0.358	0.529	0.034	0.390
0	0.5	0.4	0.419	1.000	0.427	0.599	0.052	0.383
0.2	0	0.2	0.103	1.000	0.030	0.162	0.113	0.115
0.2	0.1	0.2	0.103	1.000	0.103	0.246	0.114	0.176
0.2	0.2	0.2	0.203	1.000	0.103	0.237	0.202	0.243
0.2	0.3	0.2	0.203	1.000	0.193	0.226	0.314	0.330
0.2	0.4	0.2	0.301	1.000	0.301	0.331	0.416	0.419
0.2	0.5	0.2	0.422	1.000	0.402	0.512	0.515	0.613

Table B.3: Estimated Rejection Probabilities with $\gamma = \sigma_\mu^2 = \rho = 0$. Sample size: $N = 25, T = 10$

λ	τ	δ	RS_δ^*	RS_δ	RS_τ^*	RS_τ	RS_λ^*	RS_λ
0	0.2	0	0.021	0.999	0.907	0.991	0.051	0.997
0.1	0.2	0	0.380	0.999	0.971	0.999	0.061	1.000
0.2	0.2	0	0.051	0.996	0.991	1.000	0.166	1.000
0.3	0.2	0	0.470	0.989	0.984	1.000	0.358	1.000
0.4	0.2	0	0.028	0.965	0.961	1.000	0.639	1.000
0.5	0.2	0	0.039	0.914	0.911	1.000	0.859	1.000
0	0	0.2	0.091	1.000	0.261	0.870	0.037	0.950
0.1	0	0.2	0.112	1.000	0.394	0.968	0.122	0.991
0.2	0	0.2	0.263	1.000	0.328	0.997	0.151	0.997
0.3	0	0.2	0.309	1.000	0.479	0.999	0.127	0.998
0.4	0	0.2	0.398	1.000	0.496	1.000	0.288	1.000
0.5	0	0.2	0.402	0.998	0.593	1.000	0.578	1.000
0	0	0.4	0.116	1.000	0.070	0.977	0.052	0.987
0.1	0	0.4	0.127	1.000	0.341	0.998	0.147	1.000
0.2	0	0.4	0.320	1.000	0.418	1.000	0.149	1.000
0.3	0	0.4	0.309	1.000	0.476	1.000	0.264	1.000
0.4	0	0.4	0.419	1.000	0.489	1.000	0.153	1.000
0.5	0	0.4	0.393	1.000	0.598	1.000	0.314	1.000
0	0.2	0.2	0.115	1.000	0.897	1.000	0.070	1.000
0.1	0.2	0.2	0.109	1.000	0.957	1.000	0.116	1.000
0.2	0.2	0.2	0.271	1.000	0.979	1.000	0.243	1.000
0.3	0.2	0.2	0.314	1.000	0.990	1.000	0.199	1.000
0.4	0.2	0.2	0.519	1.000	1.000	1.000	0.241	1.000
0.5	0.2	0.2	0.602	0.999	0.998	1.000	0.485	1.000
0	0	0.2	0.081	1.000	0.086	0.881	0.056	0.950
0	0.1	0.2	0.102	1.000	0.788	0.993	0.069	0.998
0	0.2	0.2	0.363	1.000	0.862	1.000	0.071	1.000
0	0.3	0.2	0.309	1.000	0.923	1.000	0.050	1.000
0	0.4	0.2	0.378	1.000	0.944	1.000	0.080	1.000
0	0.5	0.2	0.410	1.000	0.936	1.000	0.072	1.000
0	0	0.4	0.092	1.000	0.094	0.985	0.058	0.990
0	0.1	0.4	0.202	1.000	0.822	1.000	0.028	0.999
0	0.2	0.4	0.343	1.000	0.885	1.000	0.036	1.000
0	0.3	0.4	0.409	1.000	0.915	1.000	0.041	1.000
0	0.4	0.4	0.418	1.000	0.912	1.000	0.041	1.000
0	0.5	0.4	0.471	1.000	0.932	1.000	0.047	1.000

Table B.4: Estimated Rejection Probabilities with $\gamma = \sigma_\mu^2 = \rho = 0$. Sample size: $N = 25, T = 10$

λ	τ	δ	$RS_\delta{}^*$	RS_δ	$RS_\tau{}^*$	RS_τ	$RS_\lambda{}^*$	RS_λ
0	0.2	0	0.021	0.999	0.907	0.991	0.051	0.997
0.1	0.2	0	0.380	0.999	0.971	0.999	0.061	1.000
0.2	0.2	0	0.051	0.996	0.991	1.000	0.166	1.000
0.3	0.2	0	0.470	0.989	0.984	1.000	0.358	1.000
0.4	0.2	0	0.028	0.965	0.961	1.000	0.639	1.000
0.5	0.2	0	0.039	0.914	0.911	1.000	0.859	1.000
0	0	0.2	0.091	1.000	0.261	0.870	0.037	0.950
0.1	0	0.2	0.112	1.000	0.394	0.968	0.122	0.991
0.2	0	0.2	0.263	1.000	0.328	0.997	0.151	0.997
0.3	0	0.2	0.309	1.000	0.479	0.999	0.127	0.998
0.4	0	0.2	0.398	1.000	0.496	1.000	0.288	1.000
0.5	0	0.2	0.402	0.998	0.593	1.000	0.578	1.000
0	0	0.4	0.116	1.000	0.070	0.977	0.052	0.987
0.1	0	0.4	0.127	1.000	0.341	0.998	0.147	1.000
0.2	0	0.4	0.320	1.000	0.418	1.000	0.149	1.000
0.3	0	0.4	0.309	1.000	0.476	1.000	0.264	1.000
0.4	0	0.4	0.419	1.000	0.489	1.000	0.153	1.000
0.5	0	0.4	0.393	1.000	0.598	1.000	0.314	1.000
0	0.2	0.2	0.115	1.000	0.897	1.000	0.070	1.000
0.1	0.2	0.2	0.109	1.000	0.957	1.000	0.116	1.000
0.2	0.2	0.2	0.271	1.000	0.979	1.000	0.243	1.000
0.3	0.2	0.2	0.314	1.000	0.990	1.000	0.199	1.000
0.4	0.2	0.2	0.519	1.000	1.000	1.000	0.241	1.000
0.5	0.2	0.2	0.602	0.999	0.998	1.000	0.485	1.000
0	0	0.2	0.081	1.000	0.086	0.881	0.056	0.950
0	0.1	0.2	0.102	1.000	0.788	0.993	0.069	0.998
0	0.2	0.2	0.363	1.000	0.862	1.000	0.071	1.000
0	0.3	0.2	0.309	1.000	0.923	1.000	0.050	1.000
0	0.4	0.2	0.378	1.000	0.944	1.000	0.080	1.000
0	0.5	0.2	0.410	1.000	0.936	1.000	0.072	1.000
0	0	0.4	0.092	1.000	0.094	0.985	0.058	0.990
0	0.1	0.4	0.202	1.000	0.822	1.000	0.028	0.999
0	0.2	0.4	0.343	1.000	0.885	1.000	0.036	1.000
0	0.3	0.4	0.409	1.000	0.915	1.000	0.041	1.000
0	0.4	0.4	0.418	1.000	0.912	1.000	0.041	1.000
0	0.5	0.4	0.471	1.000	0.932	1.000	0.047	1.000